Engineering Asset Management

Proceedings of the First World Congress on Engineering Asset Management (WCEAM) 2006 *Gold Coast, Queensland, Australia, 11–14 July 2006*

Joseph Mathew
Lin Ma
Andy Tan
Deryk Anderson
Editors

COMPACT disc CD-ROM

ISBN 1-84628-583-6

Springer

Proceedings of the 1st World Congress on Engineering Asset Management (WCEAM)

11 – 14 July 2006

Editors
Joseph Mathew, Jim Kennedy, Lin Ma, Andy Tan
and Deryk Anderson

Published by
Springer-Verlag London Ltd

ISBN 1-84628-583-6

Proceedings of the 1st World Congress on Engineering Asset Management (WCEAM)
Conrad Jupiters, Gold Coast, Australia
11 – 14 July 2006

PREFACE

It is with great pleasure that we welcome you to the inaugural World Congress on Engineering Asset Management (WCEAM) being held at the Conrad Jupiters Hotel on the Gold Coast from July 11 to 14, 2006. More than 170 authors from 28 countries have contributed over 160 papers to be presented over the first three days of the conference. Day four will be host to a series of workshops devoted to the practice of various aspects of Engineering Asset Management.

WCEAM is a new annual global forum on the various multidisciplinary aspects of Engineering Asset Management. It deals with the presentation and publication of outputs of research and development activities as well as the application of knowledge in the practical aspects of:

- strategic asset management
- risk management in asset management
- design and life-cycle integrity of physical assets
- asset performance and level of service models
- financial analysis methods for physical assets
- reliability modelling and prognostics
- information systems and knowledge management
- asset data management, warehousing and mining
- condition monitoring and intelligent maintenance
- intelligent sensors and devices
- regulations and standards in asset management
- human dimensions in integrated asset management
- education and training in asset management and
- performance management in asset management.

We have attracted academics, practitioners and scientists from around the world to share their knowledge in this important emerging transdiscipline that impacts on almost every aspect of daily life.

All full papers published in these proceedings have been refereed by specialist members of a peer review panel for technical merit. It should be noted that some papers may contain grammatical errors in their current published state. Given timing constraints in the publication process, the Editors elected not to edit these papers to this level of detail. Nevertheless, the Review Panel has accepted these papers based on technical merit.

The vision for the WCEAM originated from the CRC for Integrated Engineering Asset Management (CIEAM) in 2004 during free form discussions at the closing of the 2004 Asia Pacific Conference on Systems Integrity and Maintenance (ACSIM) held in New Delhi, India. It was recognised that the ACSIM forum needed to expand its scope to include the research and development outputs from this growing transdiscipline. To facilitate rapid uptake for an inaugural event and in a spirit of collaboration, several international conference organising committees were approached to combine their efforts for this 1st WCEAM. These included the International Conference on Maintenance Societies (ICOMS 2006) hosted by the Maintenance Engineering Society of Australia (MESA), the International Maintenance Systems (IMS) Conference hosted by the IMS Centre, USA, The Condition Monitoring Conference hosted by the British Institute for Non-Destructive Testing (BINDT) and ACSIM hosted by CIEAM.

WCEAM Sponsorship
We would like to gratefully acknowledge the support of **Transfield Services** as Gold Sponsor of the ICOMS conference since 1994, which is being held jointly with WCEAM, the **Queensland Government through the Department of State Development and Innovation** as Silver Sponsor, and the **Beenleigh Merrimac Pimpama Alliance** as Bronze Sponsor of 1st WCEAM.

1st WCEAM is also hosting a **state-of-the-art exhibition** providing an excellent opportunity for suppliers of equipment and services in EAM to showcase their offerings in this growing market.

Vote of Thanks

We would like to thank all the members of the WCEAM Organising Committee for the enormous effort they have contributed to making this conference a success. Thanks are also due to members of the WCEAM International Steering Committee and Industry Advisory Group for their efforts in both reviewing papers as well as promoting the congress to their contacts.

In particular we note the importance of the local asset management bodies that we team with to make these international events a success. In this case the invaluable role of MESA and its Brisbane Chapter which are manned by volunteers was essential in achieving the successful outcome that this first WCEAM has become.

In closing, we would like to invite you to enjoy 1st WCEAM on Australia's Gold Coast and we encourage you to take this opportunity to network with your colleagues from around the world. Our social events have been carefully selected to maximise opportunities for you to do so. We look forward to your participation in making this a most memorable event for you and your fellow delegates.

We trust you will also sample some of the delights of this region by taking time out after the congress to enjoy what it and the state of Queensland has to offer.

Once again, welcome, and thank you for being a part of 1st WCEAM.

Professor Joe Mathew CIEAM
WCEAM Chair

Jim Kennedy MESA
WCEAM Vice Chair

TABLE OF CONTENTS

Conference Organisers

Centre for Integrated Engineering Asset Management, www.cieam.com
Click here for video

The Maintenance Engineering Society of Australia Inc, www.mesa.org.au
Click here for video about the Excellence Awards

International Steering Committee

Chair: Professor Joe Mathew Chief Executive Officer, CIEAM. Australia
Vice-Chair: Jim Kennedy, National Chairman, MESA, Australia.

Co-Chairs: Professor Jay Lee and Professor Jun Ni, Co-Directors, Intelligent Maintenance Systems Center, USA.

Committee Members:
Associate Professor Jayantha Liyanage, University of Stavanger, Norway.
Associate Professor Peter Tse, City University of Hong Kong.
Dr Dimitris Kiritsis, Ecole Polytechnique Fédérale de Lausanne, Switzerland.
Dr Zhang Jing Bing, Singapore Institute of Manufacturing Technology, Singapore.
Professor Aly El-Shafei, Cairo University, Egypt.
Professor Anthony David Hope, Southhampton Solent University, UK.
Professor Antonio J Marques Cardoso, Universidade de Coimbra, Portugal.
Professor Basim Al-Najjar, Vaxjo University, Sweden.
Professor Bo-Suk Yang, Pukyong National University, Korea.
Professor Chris Mechefske, Queensland University of Technology, Australia
Professor Gao Jinji, Beijing Chemical University, China.
Professor Joe Amadi-Echendu, University of Pretoria, South Africa.
Professor Margot Weijnen, Delft University of Technology, The Netherlands.
Professor Menad Sidahmed, Universite de Technologie Compiègne, France.
Professor Mohd Salman Leong, University of Technology, Malaysia.
Professor N S Vyas, Indian Institute of Technology, Kanpur, India.
Professor Seppo Virtanen, Tampere University of Technology, Finland.
Professor Tadao Kawai, Osaka City University, Japan.

Review Panel

Dr Joe Amadi-Echendu
Mr Deryk Anderson
Mr Danny Azavedo
Mr Ray Beebe
Mr Peter Buckland
Prof Richard Clegg
Dr Aly El-Shafei
Dr Tony Hope
A/Prof Ian Howard
A/Prof Yousef Ibrahim
Prof Zhang Jing Bing
Prof Gao Jinji
Prof Rhys Jones
Prof Tadao Kawai
Mr Jim Kennedy
Dr Dimitris Kiritsis
Mr Peter Kohler
Prof Andy Koronios
Mr Ernst Krauss
Prof Jay Lee
Prof Mohd Salman Leong
Dr Jayantha Liyanage
A/Prof Lin Ma
Dr Anthonio Marques Cardoso
Prof Joe Mathew
A/Prof Chris Mechefske
Mr Mark Mueller
Prof Lakshmi Narasimhan
Prof Jun Ni
Mr Peter Pascoe
Mr Paul Peterson
Dr. Bob Platfoot
Mr Steven Pudney
Dr Menad Sidahmed
Mr Michael Sondalini
Mr Andrew Spence
Dr Fred Stapelberg
A/Prof Andy Tan
A/Prof Peter Tse
Dr Seppo Virtanen
Dr NS Vyas
Prof Bo-suk Yang
A/Prof Nong Zhang

Organising Committee

Mr Peter Oliver, Co-Chair, MESA.
Dr Lin Ma, Co-Chair, Queensland University of Technology.
Assoc Prof Andy Tan, Queensland University of Technology.
Mr Deryk Anderson, Oniqua.
Ms Sally Nugent, MESA.
Mr Steven Pudney, Queensland University of Technology.
Ms Cathie Wilson, CIEAM.
Mr Mark Mueller, MPM Global.
Mr Anton Koekmoer, Goodman Fielder.

List of Papers

Paper 193 J Gardulski
Application of STFT to Early Fault Detection of Automotive Shock Absorber

Paper 235 D Kiritsis HB Jun A Bufardi D Pli M Anastasiou & P Xirouchakis
Engineering Asset Management Using Product Embedded Information Device (PEID): PROMISE Approach with Application to Telecom Assets

Sponsors

Transfield Services, Gold Sponsor
agbym@transfieldservices.com
www.transfieldservices.com

Department of State Development, Trade & Innovation, Silver Session
helena.campbell@qld.gov.au
www.sdi.qld.gov.au

Beenleigh Merrimac Pimpama Alliance, Bronze Session
jessica.banks@bmpalliance.com.au
www.bmpalliance.com.au

Assetivity Pty Ltd, Poster and Lunch
sandy.dunn@assetivity.com.au
www.assetivity.com.au, www.plant

Central Queensland University, Morning Tea
j.rees@cqu.edu.au
www.cqu.edu.au

Exhibitors

3pi International Pty Ltd
maureenf@3pi.com.au
www.3pi.com.au

AP Energy Business Publications Pte Ltd
eraj@safan.com
www.safan.com

ARMS Reliability Engineers
dgloster@reliability.com.au
www.reliability.com.au

Asset Capability Management
angelo@assetcapability.com.au
www.assetcapability.com

Asset Management Council
carolynne.vandercingel@mesa.org.au
www.amcouncil.com.au

Assetivity Pty Ltd
sandy.dunn@assetivity.com.au
www.assetivity.com.au, www.plant

Beenleigh Merrimac Pimpama Alliance
jessica.banks@bmpalliance.com.au
www.bmpalliance.com.au

Capability by Design
pkohler@cbdesign.com.au
www.cpdesign.com.au

Central Queensland University
j.rees@cqu.edu.au
www.cqu.edu.au

CRC for Integrated Engineering Asset Management
k.snell@cieam.com
www.cieam.com

Department of State Development, Trade & Innovatio
helena.campbell@qld.gov.au
www.sdi.qld.gov.au

Dotmar EPP Pty Ltd
roger.bartlett@dotmar.com.au
www.dotmar.com.au

Initiate Action
pslater@initiateaction.com
www.initiateaction.com

Machine Monitor Pty Ltd
alan.gore@machinemonitor.com
www.machinemonitor.com

Mechatricity Pty Ltd
mbennett@rylson.com.au
www.rylson.com.au

Mobius Software Pty Ltd
jason@iLearnInteractive.com
www.iLearnInteractive.com

Monash University
Ray.Beebe@sci.monash.edu.au
www.gippsland.monash.edu.au/science/mre

OMCS International
steve@omcsinternational.com
www.pmo2000.com

Oniqua Pty Ltd
David.Watkins@oniqua.com
www.oniqua.com

Reliabilityweb.com
bpartipilo@reliabilityweb.com

SIRF Rt
melissa.cameron@sirfrt.com.au
www.rcart.com.au

Smart Engineering Asset Management Laboratory
meac@cityu.edu.hk

SMRP
chuck.armbruster@jmsmucker.com
www.smpr.org

Spectraquest
suri@spectraquest.com
www.spectraquest.com

Springer
Anthony.Doyle@springer.com

The Asset Partnership Pty Ltd
stephen.young@assetpartnership.com
www.assetpartnership.com

Transfield Services
agbym@transfieldservices.com
www.transfieldservices.com

UE Systems
mkgoodman@att.net

World Congress - Engineering Asset Management
stever@desmaint.com
www.desmaint.com

FAULT SELF RECOVERY OF INDUSTRIAL PLANTS -VISION AND STATE OF THE ART

GAO Jinji

Diagnosis & Self-recovery Engineering Research Center
Beijing University of Chemical Technology, Beijing 100029 China)

Abstract: Machine faults can be contained or even eliminated by the machine system itself during operation, i.e. machines with "fault self-recovery" function can be built. Engineers have always aspired to invent such machines. The scientific basis, its connotations, theory, and several applications of fault self-recovery technologies are discussed in this keynote presentation. FSR (Fault Self-recovery Regulation), RCIMS (Reliability Centered Intelligent Maintenance System) and ASSR (Abnormal Situation Self-recovery Regulation) that integrates information, diagnosis, artificial intelligence, and active control technologies, are introduced. By FSR based design, the responsibility of maintenance duty can be shifted from users to the manufacturers, and the safe operation period of machines can be prolonged. The concepts presented in this keynote are important to the development of Intelligent Maintenance Systems.

Key Words: industrial plant, fault self-recovery, fault self-recovery regulation (FSR), reliability centered intelligent maintenance system (RCIMS), abnormal situation self-recovery regulation (ASSR)

1 INTRODUCTION

Reviewing the course of evolution from single-celled lives to human society, it might be surmised that machines are a continuation of lives, in the same way, man's fighting against faults of machinery is an extension of man's fighting against diseases. Since Watt's invention of steam engine in 1781 and its widespread use in industrialization in around 1800, only some 200 years has passed. Up to now, Men fight against machine faults mainly by "curing" them, i.e., by repairing them after shutting-down. On the other hand, the diseases have been with human for more than 3 million years, but the history of real medical activity is merely about 10 thousands years. Nevertheless, human race is not vanquished or eliminated by diseases. The answer lies in self-healing mechanism and ability of life. Could a machine's fault be controlled and eliminated automatically during operation? Could fault be "healed", like disease of humans and animals? Could there be a parallel drawn between fault-fighting and disease-fighting? Could healing approach of machine, in addition to curing method, be invented and developed? Those are important questions that worth further consideration. [1]

Modern process equipment is becoming bigger, faster, more automatic and intelligent. It is especially true with those high-speed turbines, pumps, compressors, centrifuges and other major machines, which are widely used in petrochemical, electrical, metallurgical, including non-ferrous metallurgical and other process industries, and closely bonded with process to form a big system. Faults of such machines may cause serious accident and huge economic loss. Beginning in 1960s, the international technological cycle has developed condition monitoring and diagnosis technology, putting Predictive Maintenance and Intelligent Maintenance into practices among industrial enterprises. And the widespread use of emergency shutdown (ESD) system has contributed a lot to ensure safety of plants. But, fault-caused shutdown is not only a source of economic loss, but also a source of new faults.

As indicated by engineering practices, except for small proportional of abrupt incidents, most faults come into being and develop gradually. If detected early, most big faults can be avoided given that proper measures are taken. One major cause of current problem is the lack of foresight to check erroneous condition at the very beginning. Traditionally, from the designing and manufacturing stages of machines to their running or operating stage, there are very few, if any, considerations about the gradual deterioration of faults in those machines, thus the chance to control the faults is often missed and failure of the machines becomes inevitable. Another cause of such situation is that the machines are often not well-integrated with monitoring and control system, and lack of intelligence.

Emerging and developing of the Cybernetics has deepened the understanding about oneness of hylic world. The Cybernetics discloses the similarity and synergetic behaviors between organic life activities and inorganic machine operations, in the aspects of informatics, Cybernetics, and Systematics, and thus, breaks the boundaries between lives and machines. [2] Between the control of disease and that of fault, there are similarities too; therefore, the principles of medicine science can be adapted to the study of diagnosis, prediction and prevention of fault. The history of medicine science is more than 10 thousands years, while the fault diagnosis began only some 40 years ago in 1960s. The theories and precious experiences of medicine practices can be expected to guide the engineering practice of fault diagnosis and prevention.

Under the guidance of the Systematics breaking the old specialty barrier, transplanting the idea of "self-healing" therapeutic method from medicinal science, integrating the technologies of condition monitoring and diagnosis, artificial intelligence, active control, fault tolerant control, embedded technology, etc., in this paper the fault self-recovery regulation (FSR) technology of machine is discussed with the prevention and self-healing of faults as the goal. The SFR technology will provide machine with fault self-recovery capability, with which fault-caused shutdown could be reduced significantly. This paper mainly focuses on the theory of fault self-recovery and its engineering applications on machines and plant systems.

2 THE THEORY OF EQUIPMENT'S FAULT SELF-RECOVERY

Equipment is a materialization of human intelligence and an extension of human function. All equipments are designed to work orderly and steadily under certain design condition range. However, due to internal change and/or outer disturbance during operation, the equipment's actual working condition can deviate from its design condition range, consequently, making the equipment enter fault or abnormal situation. Fault self-recovery regulation is a method of retaining or restoring the equipment's working condition within its original design condition range using the technologies of intelligent diagnosis and prediction and active control during operation so as to ensure steady and orderly running of the equipment.

Information can reveal the change of a structural system, its inputs and outputs and their interrelations. When there is an abnormal change of the interrelations and inerreactions among machine's innner sub-systems, or an abnormal change of the interrelations and inerreactions between these sub-systems and their mass inputs/outputs, energy inputs/outputs and environments, i.e. when the machine deviates from orderly condition and enters a pre-fault state, information can be detected and processed realtimely to indicates fault syptoms, then diagnosis and prediction can be made.

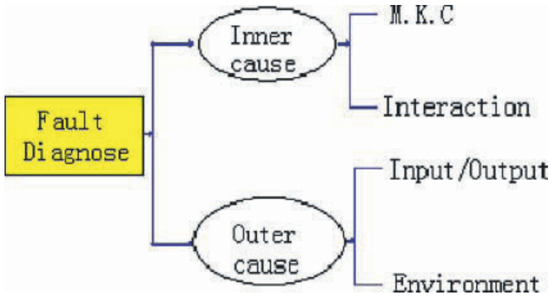


Fig. 1 Fault diagnosis both in inner and outer causes

2.1 Diagnoses and Prediction of Equipment's Faults

Following the idea of the Systematics, the pattern of fault origination and development can be studied. As shown in Fig.1, in the study, both information and relations should be taken into account, and both inner causes (i.e. change of mass, rigidity, damp etc. and their inner-reciprocity) and outer causes (input/output of mass and energy, and the influence of environment) should be considered. This will make possible early detection and diagnosis of initial cause of the fault of complex equipment system. Besides, by investigating the cause of a fault, the condition which may cause a fault can be predicted. In time domain and space domain, the evolution from failure analysis to initial cause tracking and to prediction of fault-causing condition is shown in Fig.2.

Diagnosis means the study of a complex equipment system, it's input/output of mass and energy and its interaction with environment. Through analyses of the monitoring information, the interactions and interrelations can be studied, and the rule of fault creation, development and prevention can be found, and the initial cause can be tracked through the use of fault causal chain.

Prediction has two meanings: to predict the developing trend of current fault; and to study the conditions that might lead to fault. It would be necessary to process and analyze diverse fault symptoms, to extract feature parameters, to evaluate the concordance of mass and energy input/output, to check the abnormal change of discharged material, to calculate and analyze the non-linear relation between the state of equipment system and the state parameters, so that the causes of a fault, especially the preconditions and features of a fault can be detected and a rational prediction can be made when the fault is at the beginning state.

2.2 The Theory of Fault Self-Recovery

The definition of fault self-recovery regulation: By real-time monitoring and analyzing the possibly fault-causing conditions and early fault symptoms during operation of an equipment system, through diagnosis and prediction, intelligent decision making and active control, the preconditions of a fault can be eliminated, or the fault can be automatically terminated at its beginning stage, or the possible fault can be found and intercepted, and then eliminated even before it really forms. [3]

In the modeling process of fault self-recovery system, the promptness and correctness of system response has to be studied. Since the target system is a complex system with a considerable amount of information, it is very important to extract useful information from large amount of information and to take proper measures timely, so as to fulfill self-recovery regulation. The system analysis method of the system engineering theory might be used to extract and analyze systemic information and to build an intelligent expert knowledge database to provide useful control strategy to assist intelligent decision-making.

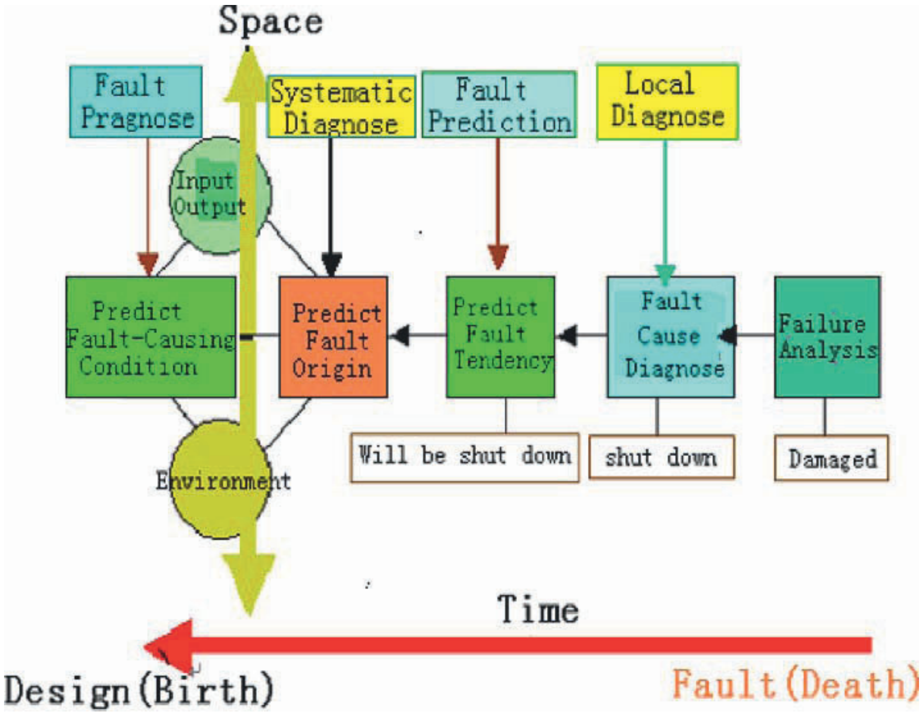


Fig.2 The evolution from failure analysis to fault-causing condition predict

The control system that aims at fault control and elimination, i.e., a fault self-recovery system comprises of real-time monitoring unit, comparison and judgment unit, diagnosis and prediction unit, intelligent decision-making unit, inter-transmitting link and execution unit. [4] The general style of fault self-recovery system is shown in Figure 3.

To build a fault self-recovery mechanism through the use of active control for the purpose of fault prevention and elimination, the major forms of fault self-recovery include:

1) The main idea of FSR is to change the working condition and state parameters. It can be done through following methods:

Compensation of disturbance. Activate a counter-reaction to prevent the creation and development of fault.

Removal of disturbance. Eliminate the fault-causing disturbance to eliminate a fault.

Deviation adjustment. Eliminate the fault-causing condition to prevent fault from happening and developing.

Hybrid approach. A combination of the above.

2) Sometimes, it might be possible to change allowable working range through dynamic regulation. The working range can be a dot, line of a multi-dimensioned region.

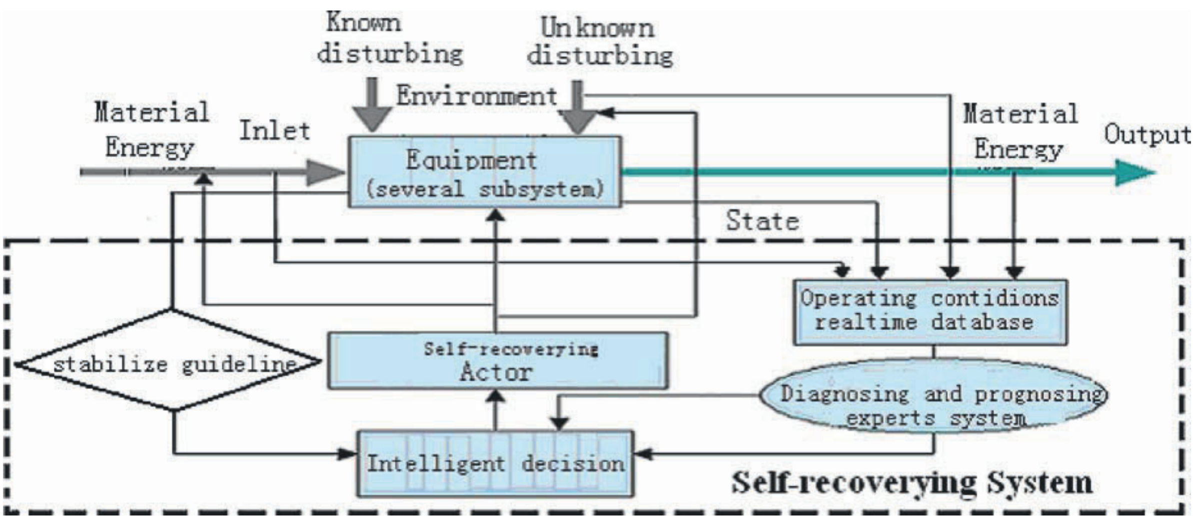


Fig.3 Model of fault self-recovering

3 ENGINEERING APPLICATIONS OF FAULT SELF-RECOVERY REGULATION OF EQUIPMENT

Above mentioned study can provide scientific basis and core technology for the development of industrial fault self-recovery regulation (FSR) system. Once a fault self-recovery regulation (FSR) system is installed before emergency shutdown(ESD) system, which is widely used in process industry, the automatic emergency shutdown cause by interlink device can be greatly reduced without necessarily having to sacrifice safety. And based on it, a new generation of equipment~~s~~, which is capable of self-healing, can be produced, changing or reducing the traditional maintenance works which mainly rely human intervene. In Fig. 4 it is illustrated how a FSR system bridges the gaps among state monitoring, diagnosis, maintenance of machine and process and process control, thus to build process machines integrated with diagnosis, supervisory and control system, consequently to achieve fault auto-maintenance, i.e., self-recovery function.

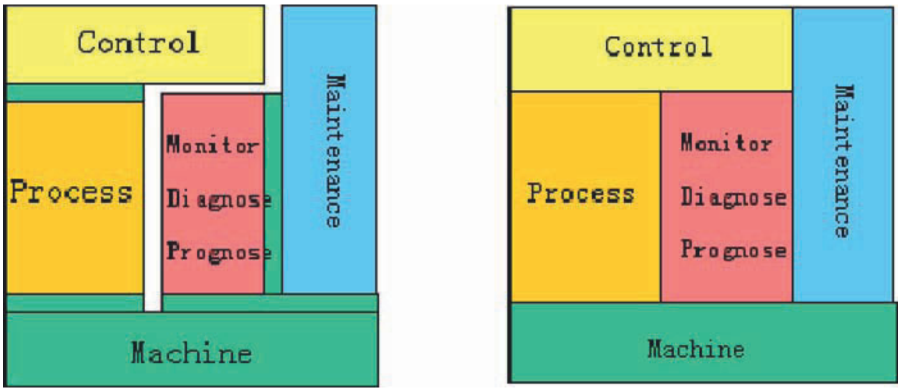


Fig. 4 Integration of process, machine, control and maintenance

Typical example applications of fault self-recovery in centrifugal compressor include:

Automatic balance of rotor: Automatic balancing system for high-speed rotor can be applied to single rotor, including vertical rotors and cantilever rotors, and multi-rotor systems (Form (a) of fault self-recovery: compensation of disturbance);

Active magnetic bearing: Diagnosis and self-recovery regulation using active magnetic bearing to deal with excess vibration (Form (d) of fault self-recovery: hybrid approach) ;

Injection of de-coking water: Under some situation, de-mineral water can be injected into the case of centrifugal compressor to reduce temperature thus to prevent coking on the rotor blades (Form (c) of fault self-recovery: adjustment to deviation);

De-coking through oil flush: washing the coke away from the blades of rotor by injecting oil into the case of centrifugal compressor (From (b) of fault self-recovery: removal of disturbance);

Adjustment of lube oil temperature for vibration control: By Adjusting pressure and temperature of lube oil, the stiffness and damp of bearing oil film can be changed to suppress vibration, such as 1/2 frequency (or low frequency) unstable vibration, to realize fault self-recovery (Form (c) of fault self-recovery: adjustment to deviation);

Anti-surge system: Through automatic reflux of feed, surge due to low suction flow can be prevented. (Form (c) of fault self-recovery: deviation adjustment);

Axial displacement self-recovery regulation: For high-speed turbo machine, through self-recovery regulation of axial displacement and adjustment of oil pressure to thrust bearing to prevent thrust bearing damage.(Form (c) of fault self-recovery: deviation adjustment);

Condition-based dry gas seal flushing: By monitoring whether the gas film carries liquid, and if so, blowing in some clean gas, the seal damage caused by loss of gas film stability can be avoided (Form (b) of fault self-recovery: removal of disturbance).

The following is a brief introduction to rotor auto-balance and axial displacement self-recovery regulation technology.

3.1 Auto-balancer

A large portion of faults in a rotating machine comes from rotor imbalance. This is especially true with the imbalance produced during operation due to coking, dirt forming or bending of rotor caused by thermal expansion. Such an imbalance can cause very severe vibration problem. Manual dynamic balancing is traditionally used to deal with such faults. If an online automatic balancing system is used, the imbalance of rotor or shaft can be automatically eliminated, and the fault can be recovered without having to shutdown the machine. The structure of auto-balancer for high-speed rotating machine is shown in Fig. 5. Self-recovery controller analyses oscillation signal from sensors and decides if the vibration is caused by rotor imbalance. If so, it calculates the magnitude and phase of the imbalance and sends a command to stationary ring to adjust the mass distribution of rotating ring to produce a counter-poise to original imbalance, thus to reduce vibration. The balancing heads often used in engineering are magnetic and hydraulic types. [5]

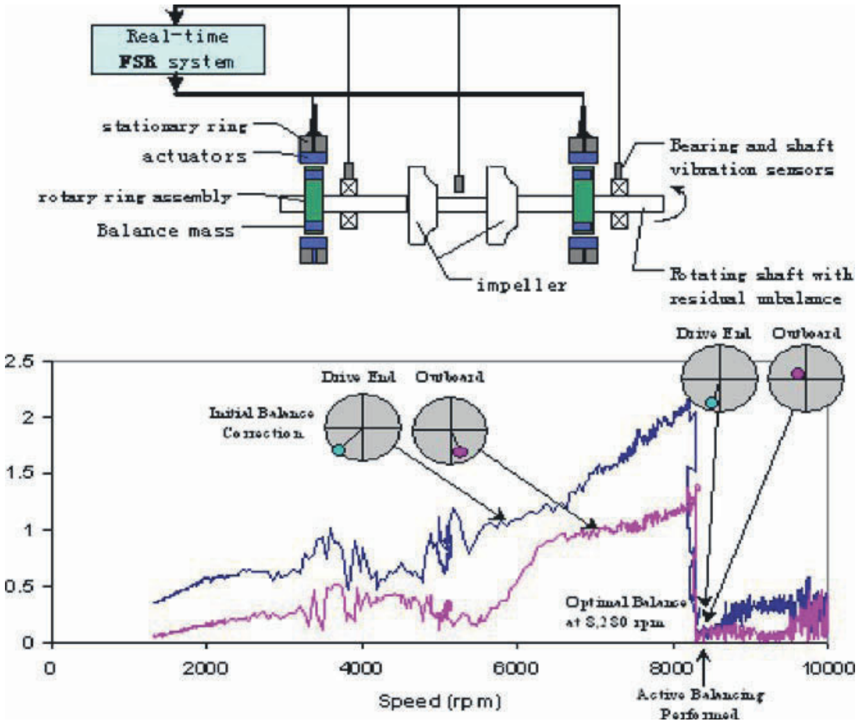
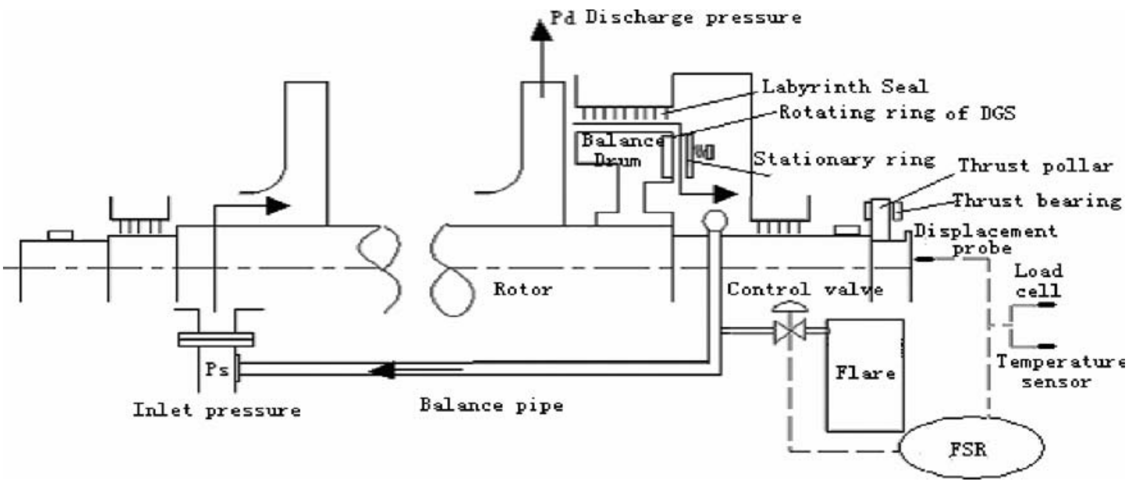


Fig.5 Principle of Auto-balancer

3.2 Fault self-recovery regulation of axial displacement

During the operation of rotary fluid machines (i.e. pumps, fans, centrifugal compressors), the interaction between rotor and fluid may produce an dynamic axial force. Often, most of such force is balanced by a pressure-difference type balancing drum, which reduces a large proportion of the dynamic axial force to reduce a load of thrust bearing. In abnormal situation machine must be shutdown by ESD system to prevent trust bearing damage from exceeding the thrust bearing's load.



Now under development a new kind of axial thrust balancing device with fault self-recovery regulation of axial displacement comprises a thrust detecting unit, a controller and a fluid flow rate adjusting unit to form a close control loop. The thrust detecting unit directly or indirectly measures the change of axial force applied to the thrust bearing, then the controller accordingly controls the flow rate adjusting unit to change the pressure difference between the balancing drum's two sides, in order to automatically adjust the bearing thrust load, therefore to prevent the thrust bearing damage from exceeding the thrust bearing's load, and in the same time to avoid unplanned trip and shutdown. Fig.6 and 7 illustrates the principle of fault self-recovery regulation of axial displacement on a high-speed turbine machine.

Since centrifugal compressors are often key equipments in petrochemical, metallurgical, and other process industries, and often there is no spare compressor. By adding a FSR system before ESD, the unplanned shutdown can be reduced without sacrifice safety. That is of obvious economical benefit.

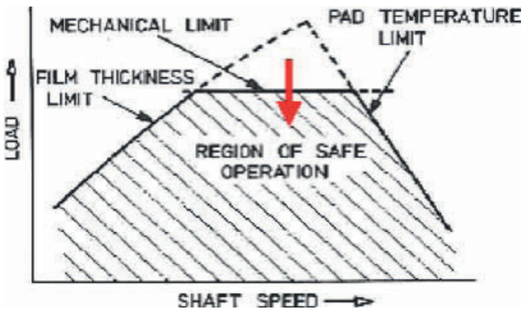


Fig.7 Safe operation region of thrust bearing

4 SELF-RECOVERY MECHANISM BASED INTELLIGENT MAINTENANCE SYSTEMS FOR MACHINE GROUP

The idea of fault self-recovery can be extended to cover the whole complex system of pump/machine group, including human, i.e., a man-machine-process-environment system. Gathering the knowledge of experts group and unitizing the power of computers, an operable intelligent platform that solves the problems of the man-machine-process-environment system can be implemented, and reliability centered intelligent maintenance system (RCIMS) can be built. See figure 8 and 9.

A networked reliability centered intelligent maintenance system (RCIMS) bases on monitoring and diagnosis system. It structures user, maintenance plan management, spare part management and maintenance groups to form a self-organizing (Self-healing) mechanism. An engineering demonstration of networked RCIMS has been built using system engineering approach to manage the moving equipments of Liaoyang petro-chemical plant, China National Petroleum Co. see Fig 10.

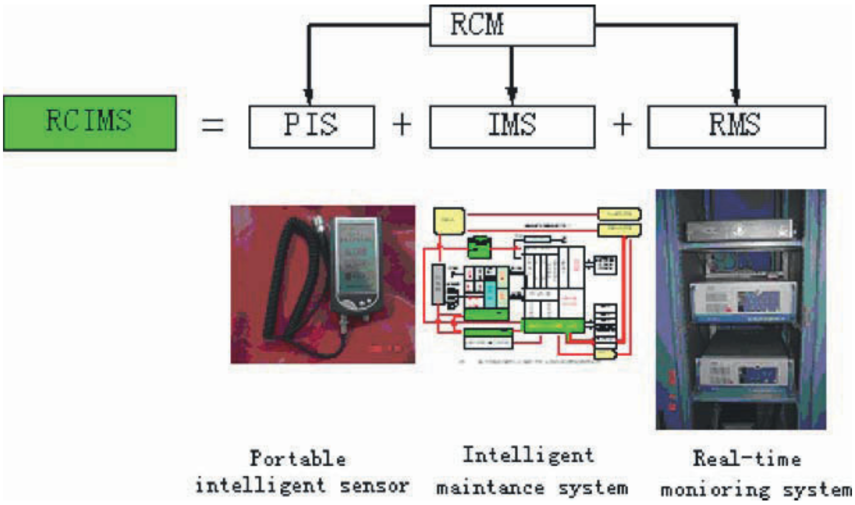


Fig.8 Reliability centered networking intelligent maintenance system

5 FAULT SELF-RECOVERY REGULATION OF ABNORMAL SITUATION IN PROCESS PRODUCTION UNIT

The idea of FSR can be used to prevent a man-machine-process-environment complex system from entering an abnormal situation. Such a system can be called an ASSR system (Abnormal Situation Self-recovery Regulation). Abnormal situation means an abnormal state of a plant that comes into being when severe disturbance exceeding design limit, which cannot be properly handled by control system. As some investigations stated, in the U.S. alone, the accidents stem from abnormal situation cost the petrol-chemical industry up to 20 billion dollars yearly. The ratio between the direct loss and the indirect loss of accidents is about 1:10. The loss caused by abnormal situation accounts for about 3-8% of the total cost.

The timely identification, diagnosis, assessment and adjustment conducted by a FSR system can enable a plant to restore normal situation or re-enter a safe state. The general technical strategy and solution is the all-rounded and all time-cycled monitoring of abnormal situation, with the focus mainly on the study of abnormal situation detection, prediction and its self-recovery regulation.

Detection and prediction: investigation in the cause of abnormal situation and in prediction of conditions that leads to abnormal situation.

Self-recovery regulation: Study concerning how to prevent, control or eliminate the inner change or outer disturbance of a system during operation, so as to prevent the system from exceeding design limit and entering abnormal situation.

For dangerous and critical operation, an automatic device or programmed control device, interlock device or linkage device should be provided to prevent disoperation.

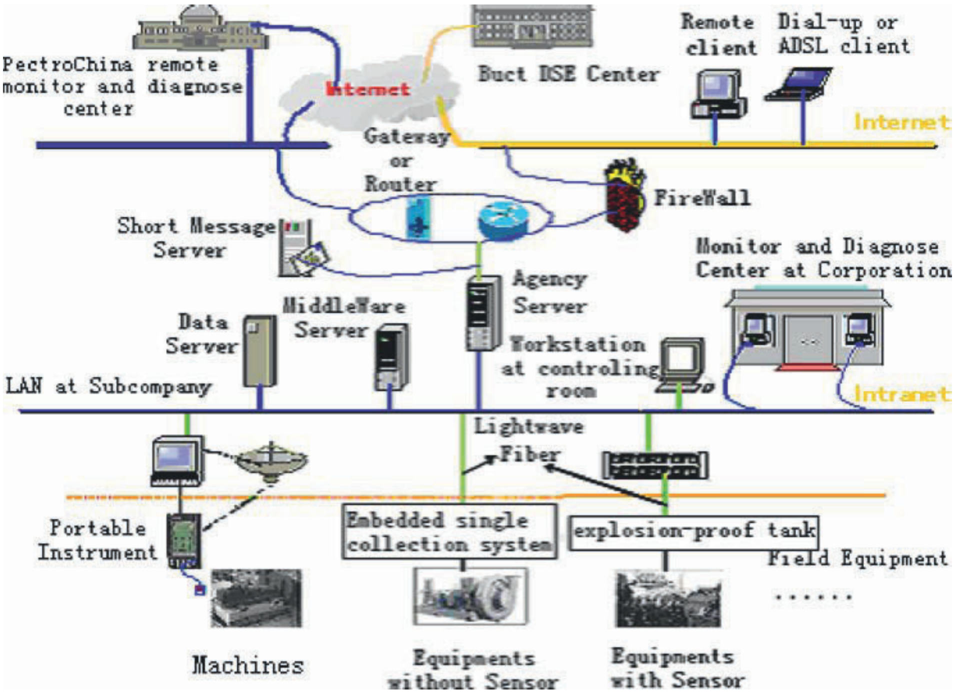


Fig.9 Network topology of reliability centered networking intelligent maintenance system

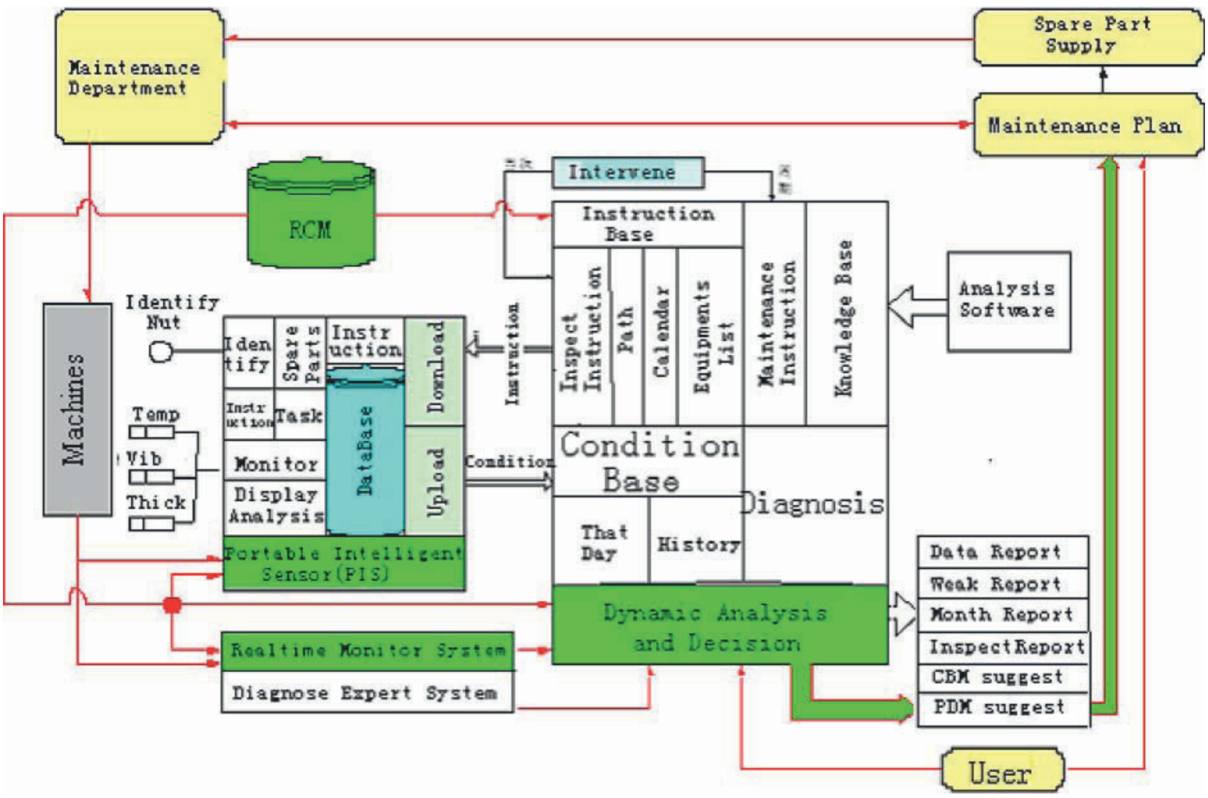


Fig10 FSR based and reliability centered networking intelligent maintenance system

6 INFORMATIONALIZED INTELLIGENT SAFETY ENSURING SYSTEM FOR PROCESS PRODUCTION UNIT

The principle of fault self-recovery and its engineering application in equipment, pump/machine groups and abnormal situation of complex process plants can effectively prevent and eliminate various kinds of faults in all time level, and thus can greatly boost safety and reliability of the process industry operations as shown in Fig. 9. [6] Following seven reductions (7 Rs) can be achieved:

1) Reduction of faults and accidents

2) Reduction of production loss due to shutdown

3) Reduction of "surplus inspection"

4) Reduction of "surplus maintenance"

5) Reduction of pollutant discharge

6) Reduction of spare part inventory

7) Reduction of equipment's dependence upon human.

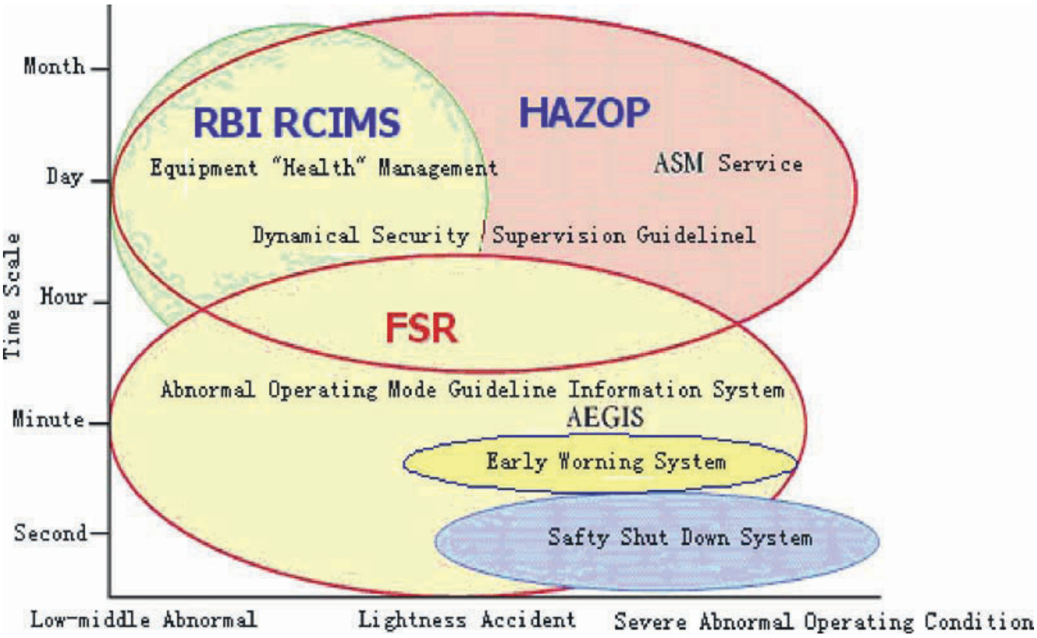


Fig.11 Principle of abnormal situation monitor and control with FSR

The FSR based design can shift the responsibility of maintenance duty from users to the manufacturers of equipment, and self-repair can shift the responsibility of remanufacture to operation which can prolong the safe operation period of equipment as shown Fig. 12.

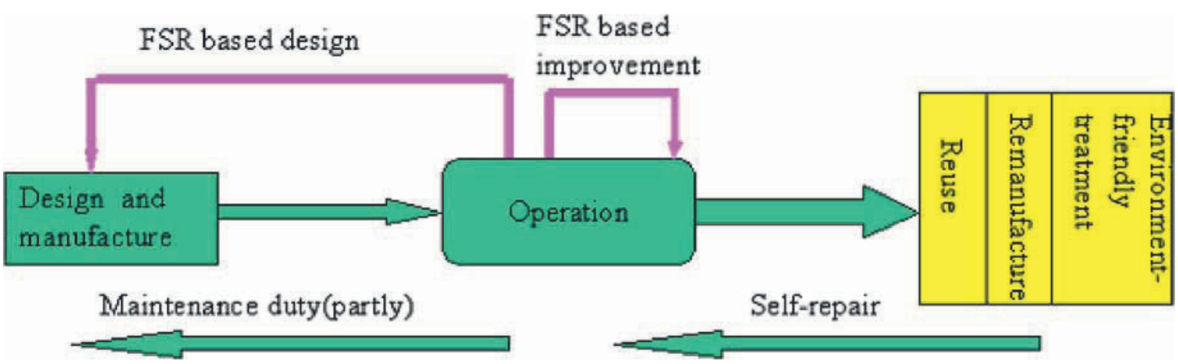


Fig.12 FSR based design can shift the responsibility of maintenance duty from users to the manufacturers

7 CONCLUSION

The fittest will survive in natural selection. From the viewpoint of scientific development, human should not only in harmony with nature, but also in harmony with their creation. The FSR can make engineers' dream come true. It will make the future equipments and production units more adapted to human demands, more harmonic with human. FSR is an advanced multidisciplinary technology, and is system engineering involving multiple personnel groups of different specialties; it opens a new horizon for Independent Innovation and has a great development potential in the future.

1) In the future, FSR can reduce the machines dependence upon human and liberate human from much of mental labors.
2) FSR can bring fighting against faults to the earlier designing stage and shift the duty of maintenance from user to the designer and manufacture;
3) FSR is an engineering method that can reduce disaster and boost productivity through fault prevention and reduction of maintenance, is very important to the long-cycle safe operation of process.
4) FSR provides a scientific base for the development of new generation equipments and process units that are maintenance free or require very little maintenance, to implement an environmental friendly and conservation-minded production process.

8 REFERENCES

[1] [Gao Jinji Think about Future Plant Medicine Engineering, Engineering Science, 2003, vol.5(12), 30~35

[2] Shen Lugeng General Systematics Beijing Broadcast institute, 2000.1

[3] Gao Jinji. A study of the fault self-recovering regulation for process equipment[A], proceedings of IMS2003, Xi'an, China 2003

[4] Gao Jinji, Jiang Zhinong. Research on fault self-recovery engineering, IMS2004 proceedings s2-c, Arles France 2004 7

[5] Shiyu Zhou, Jianjun Shi. Active Balancing and Vibration Control of Rotating Machinery: A Survey The Sock and Vibration Digest. Vol. 33. No4. July 2001 361-371 c Sage Publications

[6] Understanding and Enhancing Abnormal Situation Management in Process Control Environments: ASM Program Overview, Dr. Ted Cochran ,Dr. Peter Bullemer, Honeywell Technology Center, Minneapolis, MN

ZERO-BREAKDOWN MACHINES AND SYSTEMS: PRODUCTIVITY NEEDS FOR NEXT-GENERATION MAINTENANCE

J Lee [a] & LW Scott [b]

[a] *Ohio Eminent Scholar and Alter Chair, University of Cincinnati, OHIO, USA*

[b] *NSF Center for Intelligent Maintenance Systems (IMS) Univ. of CincinnatiCincinnati, USA* www.imscenter.net

Abstract: Predictive maintenance (or prognostics) is critical to any engineering systems to avoid system breakdown, in particular, for complex systems. With the recent advances on pervasive computing, prognostics can be easily embedded in any devices and systems. When smart machines are networked and remotely monitored, and when their data is modeled and continually analyzed with sophisticated embedded systems, it is possible to go beyond mere "predictive maintenance" to intelligent "prognostics", the process of pinpointing exactly which components of a machine are likely to fail, and when and autonomously trigger service and order spare parts. This paper addresses the paradigm shift in modern maintenance systems from traditional "fail and fix" practices to "predict and prevent" methodology. Recent advances on prognostic technologies and tools are discussed. In addition, case studies are used to validate f these developed technologies and tools.

Key Words: diagnostics, prognostics, machine performance assessment, predictive maintenance

1 INTRODUCTION

For years, maintenance has been treated as a dirty, boring and ad hoc job. It's critical for productivity but isn't yet recognized as a key component of revenue generation. The simple question is often, "Why do we need to maintain things regularly?" The answer is, "To keep things as reliable as possible." But the real question is, "How much change or degradation has occurred since the last round of maintenance?" The answer is, "I don't know." Condition-based maintenance (CBM) deals with online data. Machine conditions are constantly monitored and their signatures evaluated. However, this is done at the machine level -- one machine at a time. It's a "fail-and-fix approach." Troubleshooting is the primary purpose. Today, CBM focuses on sensors and communications. All products and machines are networked by some means. It's hard to know, though, what to do with all this data. We need to turn data into information by using computational tools to process data locally. The maintenance world of tomorrow is an information world for feature-based monitoring. Information should represent a trend, not just a status. It should offer priorities, not just show "how much." If we do that, then our productivity can be focused on asset-level utilization, not just production rates.

Today, machine field services depend on sensor-driven management systems that provide alerts, alarms and indicators. The moment the alarm sounds, it's already too late to prevent the failure. Most factory downtime is caused by these unexpected situations. There is no alert provided that looks at degradation over time. If we can monitor degradation, then we can forecast upcoming situations and perform maintenance tasks when necessary (not too early or too late). Using these techniques, maintenance can be scheduled prior to failure. With the advent of modern computing and communication technologies, more products and machines are embedded with sensors and connected through tethered-and tether-free networks. In addition, with the seamlessly integrated network systems in today's global business environment, machines and factories are networked and information and decisions are synchronized, so any downtime in the system can jeopardize the entire enterprise productivity. On the another hand, constraints on security and communication bandwidth limitations require re-engineering to minimize or mitigate the risk of data exposure.

Intelligent maintenance systems (IMS) predict and forecast equipment performance so "near-zero breakdown" status is achievable. There are two reasons for failure: equipment performance and operator/human error. Near-zero downtime focuses on predictive techniques to minimize failures. It focuses on features of machine performance. Recently, many sophisticated sensors and computerized components have been developed, which are capable of collecting data carrying the degradation information of the machine's health status and performance. The problem is that most of this data has not been fully utilized in practical use. This is mainly because the available data are not rendered in a useable form. Therefore, advanced technologies are needed to process the data and extract the useful degradation information to be utilized in maintenance decision making in a more intelligent manner. Moreover, through the integrated life-cycle management, such degradation information can be used to make improvements in every aspect of product life-cycle. World-class companies already have taken a game-changing approach, implementing a new service business model to transform maintenance systems into smart service and asset management solutions. They reduce downtime and provide the ability to look ahead at the quality of products before they ship

by closely watching equipment performance and machine degradation. Rather than reactive maintenance — "fail-and-fix" — companies can indeed move to "predict-and-prevent" maintenance.
In this paper, we first review the state-of-the-art on maintenance technologies and introduce the concept of intelligent maintenance systems using product degradation information. Afterwards, the prognostic tools recently developed will be addressed, and several case studies are provided to demonstrate these tools in practical use. These real examples prove the effectiveness and potentials of these tools, as well as address needs in future research directions.

2 STATE-OF-THE-ART REVIEWS ON MAINTENANCE TECHNOLOGIES

2.1 Maintenance Technologies Overview

Looking back on the development history and forecasting the development tendency of maintenance technologies, the roadmap to excellence in maintenance can be illustrated as in figure 1.

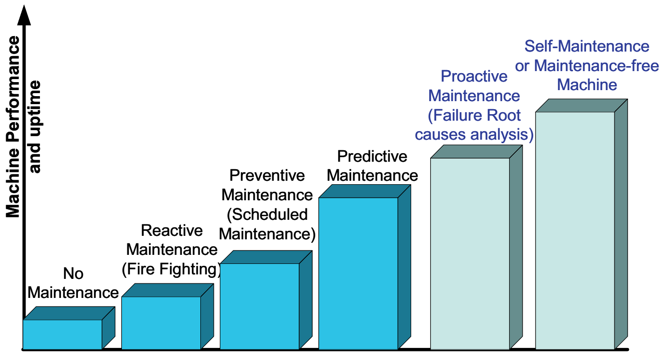


Figure 1. The development of maintenance technologies

2.1.1 No Maintenance

There are two kinds of situations in which no maintenance will occur.

- No way to fix it: the maintenance technique is not available for some special application, or the maintenance technique is not well developed at the early stage.
- Isn't worth it to fix it: some machines were designed to be used only once. Comparing with maintenance cost, it might be more cost-effective just to discard it.

Neither of the scenarios above is within the scope of the discussion here.

2.1.2 Reactive Maintenance

In plain English, the aim of reactive maintenance is just to "fix it after it's broken," since most of the time a machine brakes down without warning and it is urgent for the maintenance crew to put it back to work: this is also referred to as "fire fighting".

The reason that reactive maintenance happens is because some operations have developed through the years with very little attention given to the proper care of the machinery involved. Essentially, little to no maintenance is conducted and the machinery operates until a failure occurs. At this time, appropriate personnel are contacted to assess the situation and make the repairs as expeditiously as possible. In a situation where the damage to equipment is not a critical factor, as plenty of downtime is available, and the values of the assets are not a concern, the firefighting mode may prove to be an acceptable option. Of course, one must consider the additional cost of making repairs on an emergency basis since soliciting bids to obtain reasonable costs may not be applicable in these situations. Due to market competition and environmental/safety issues, the trend is toward appropriating an organized and efficient maintenance program as opposed to firefighting.

2.1.3 Preventive Maintenance

Preventive Maintenance (PM) is an equipment maintenance strategy based on replacing, overhauling or remanufacturing an item at a fixed interval, regardless of its condition at the time. Scheduled restoration tasks and scheduled discard tasks are both examples of Preventive Maintenance tasks.

2.1.4 Predictive Maintenance

Predictive Maintenance (PdM) is a right-on-time maintenance strategy. Predictive maintenance may be best described as a process which requires technologies and people skills, while combining and using all available diagnostic and performance data, maintenance histories, operator logs and design data to make timely decisions about maintenance requirements of major/critical equipment. It is the integration of various data, information and processes that leads to the success of a PdM program. It analyzes the trend of measured physical parameters against known engineering limits for the purpose of detecting, analyzing and correcting a problem before a failure occurs. A maintenance plan is made based on the prediction results derived from condition based monitoring. This can cost more up front than PM because of the additional monitoring hardware and software investing, manning, tooling, and education required to establish a Predictive Maintenance program. However, it offers increased equipment reliability and a sufficient advance in information to improve planning, thereby reducing unexpected downtime and operating costs.

2.1.5 Proactive Maintenance

Proactive Maintenance, in general terms, encompasses any tasks used to predict or prevent equipment failure. To be more specific, there are two working directions.

- Change the failure reactive to failure proactive by avoiding the underlying conditions that lead to machine faults and degradation. Proactive Maintenance focuses on analyzing the root cause, and not just the symptoms. It seeks to prevent or to fix the failure from the source after it identifies the root cause.
- Feed the maintenance information back to the design and operation department. Failure prevention should also be conducted in the design and operation department. The maintenance crew's job is not only to fix a machine or change parts, but they should also help by suggesting how to improve a machine's design and operation so that the failures are prevented proactively.

There is still some debate about the efficiency and failure response speed of proactive maintenance, but there is no doubt that there has been a lack of communication between maintenance and design

2.1.6 Self maintenance

Self-maintenance is a new design and system methodology. A self-maintaining machine can monitor and diagnose itself, and if any kind of failure or degradation happen, it can still maintain its functions for a while. A Self-maintaining machine doesn't belong in the conventional physical maintenance concept, but in the functional maintenance concept instead. Functional maintenance aims to recover the required function of a depredating machine by trading off functions, whereas traditional repair (physical maintenance) aims to recover the initial physical state by replacing faulty components, cleaning, etc. The way to fulfill the self-maintenance function is by adding intelligence to the machine, making it clever enough for functional maintenance. In other words, self-maintainability would be appended to an existing machine as an additional embedded reasoning system.

Another system approach to creating the self-maintaining ability is to add the self-service trigger function to a machine. The machine will then self-monitor, self-diagnose and self-trigger the service request with detailed and clear maintenance requirements. The maintenance task is still conducted by a maintenance crew, but the no gap integration of machine, maintenance schedule, dispatch system and inventory management system will minimize maintenance costs to the greatest extent and raise customer satisfaction to the highest level.

2.2 State- of-the-art review on Condition Based Maintenance (CBM)and Prognostics Technologies

Traditional reliability prediction is made based on failure time data; meanwhile, maintenance actions are initiated in terms of a reliability index. Usually such preventive maintenance schemes as block-replacement and age-replacement are time-based without considering the current health state of the product, thus are inefficient and less valuable for a customer whose individual asset is of the most concern.

The major role of degradation analysis is to investigate the evolution of physical characteristics or performance measures of a product leading to failure. A maintenance scheme, referred to as Condition Based Maintenance (CBM), is developed by considering current degradation and its evolution. CBM methods and practices have been continuously improved for the last decades. For example, a major manufacturer of elevators for high rising buildings continuously monitors the braking systems, and acceleration and deceleration of elevators globally to meet the high safety requirement. The main idea of CBM is to utilize the product degradation information extracted and identified from on-line sensing techniques to minimize the system downtime by balancing the risk of failure and achievable profits. The decision-making in CBM will focus on predictive maintenance. To do so, many diagnostic tools and methods have been developed with good successes. Sensor fusion techniques are also commonly engaged due to the inherent superiority in taking advantage of mutual information from multiple sensors (Roemer *et al*., 2001; Hansen *et al*., 1994; Reichard *et al*., 2000). For example, vibration signature analysis and oil analysis, because of their excellent capability of describing machine performance, have been successfully employed for prognostics for a long time (Kemerait, 1987; Wilson *et al*., 1999). In the mean time, alternative approaches in using time/stress, temperature, acoustic emissions (Goodenow *et al*., 2000), and ultrasonic are widely employed as well. However, major breakthrough has not been made in the development on prognostic since many methods were still based on traditional signal processing methods (Hardman *et al*., 2000).

In the recent technical literatures, a large variety of prognostic applications have been reported, and many of these were more application specific rather than generic. Some research activities about infrastructure of prognostics were also reported. From 1998 to 2000, the U.S. army logistics integration agency funded a project entitled, "Prognostics Framework" (Su *et al*., 1999). The project aimed to provide an overall architecture to manage the information provided by the individual prognostic techniques. This prognostics framework is generic and scalable, with open architecture and horizontal technology, which was intended to integrate with embedded diagnostics to provide a "total health management" capability.

Generally speaking, current prognostic approaches can be classified into three basic groups: *model-based prognostics*, *data driven prognostics*, and *hybrid prognostics*. An example of a characteristic *model-based prognostics* application includes collecting data from model-based simulations under normal and degraded conditions. Prognostic models are constructed based on different random load conditions, or modes. However, in the absence of a reliable or accurate system model, another approach to determine the remaining useful life is to monitor the trajectory of a developing fault and predict the amount of time until the developing fault reaches a predetermined level requiring action, which is the so called data driven prognostic method. An example of a hybrid method, which fuses the model-based information and sensor-based information and takes advantage of both model-driven and data-driven methods, was proposed by Hansen *et al*. (1994), by which a more reliable and accurate prognostic result can be generated.

Prognostic information obtained through intelligence embedded into the manufacturing process or equipment can also be used to improve the manufacturing and maintenance operations in order to increase process reliability and improve product quality. For instance, the ability to increase reliability of manufacturing facilities using the awareness of the deterioration levels of manufacturing equipment has been demonstrated through an example of improving robot reliability (Yamada and Takata, 2002). Moreover, a Life Cycle Unit (LCU) (Seliger *et al*., 2002) was proposed to collect usage information about key product components, enabling one to assess product reusability and facilitating reuse of products that have significant remaining useful life.

In spite of these progresses, many fundamental issues still remain:

1. Most of the developed prognostics approaches are application or equipment specific. A generic and scalable prognostic methodology or toolbox doesn't exist;
2. Currently, methods are generally focused on solving the failure prediction problem. Tools for system performance assessment and degradation prediction has not been well addressed;
3. Features used for prognostics need to be further developed;
4. Many developed prediction algorithms have been demonstrated in a laboratory environment without industrial validations.

To address these issues, a toolbox of algorithms for multi-sensor performance assessment and prediction, named Watchdog Agent™, has been developed at the Center for Intelligent Maintenance Systems (IMS) which is a multi-campus NSF Industry/University Cooperative Research Center between the University of Cincinnati and the University of Michigan. These tools enable one to quantitatively assess and predict performance degradation levels of key product components (Casoetto *et al*., 2003; Djurdjanovic *et al*., 2000; Lee, 1995; Lee, 1996), thus offering a possibility to physically realize a closed-loop product life cycle monitoring and management.

3 CURRENT DEGRADATION FEATURE EXTRACTION AND PREDICTIVE TOOLS

Watchdog Agent™ developed by the Center for IMS consists of embedded computational prognostic algorithms and a software toolbox for predicting degradation of devices and systems. The Watchdog Agent™ bases its degradation assessment on the readings from multiple sensors that measure critical properties of the process, or machinery that is being considered. It is expected that the degradation process will alter the sensor readings that are being fed into the Watchdog Agent™, and thus enable it to assess and quantify the degradation by quantitatively describing the corresponding change of sensor signatures. In addition, a model of the process or piece of equipment that is being considered, or available application specific knowledge can be used to aid the degradation process description, provided that such a model and/or such knowledge exist. The prognostic function of the watchdog is realized through trending and statistical modeling of the observed process performance signatures and/or model parameters. This allows one to predict the future behavior of these patterns and thus forecast the behavior of the process, or piece of machinery that is being considered.

In order to facilitate the use of Watchdog Agent™ in a wide variety of applications, with various requirements and limitations regarding the character of signals, available processing power, memory and storage capabilities, limited space, power consumption, personal user's preference etc, the performance assessment module of the Watchdog Agent™ has been realized in the form of a modular open architecture toolbox. The toolbox consists of different prognostics tools, including neural network based, time-series based, wavelet-based and hybrid joint time-frequency methods, etc., for predicting the degradation or performance loss on devices, process, and systems. Open architecture of the toolbox allows one to easily add new solutions to the performance assessment modules as well as to easily interchange different tools, depending on the application needs. The tools currently realizing the signal processing, feature extraction, sensor fusion and performance evaluation modules of the performance assessment Watchdog Agent™ functionality are summarized in Figure 2.

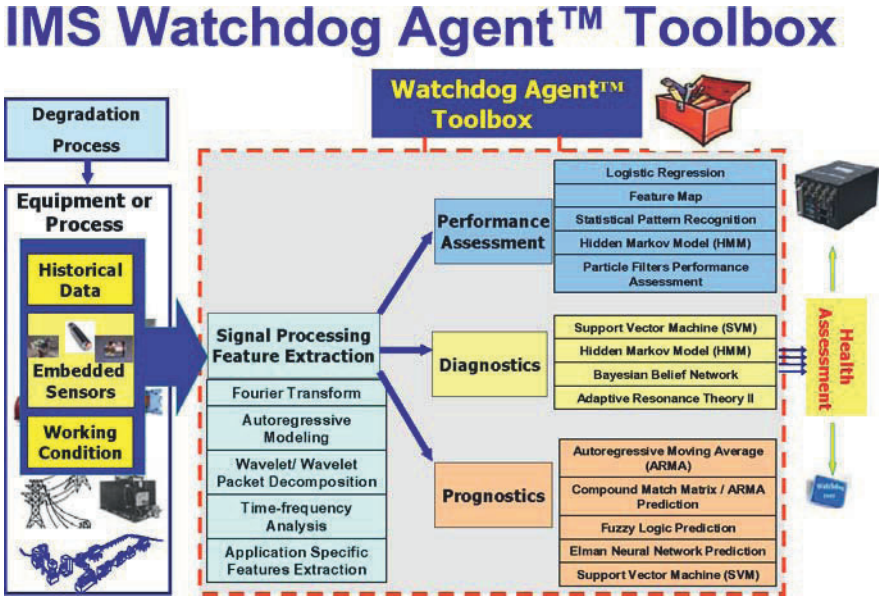


Figure 2. Watchdog Agent Prognostics Toolbox

Each of these modules is realized in several different ways to facilitate the use of Watchdog Agent™ in a wide variety of products and applications. Specifically, these modules are:

- **Sensory processing module** transforms sensor signals into domains that are the most informative of product's performance. Time-series analysis (Pandit and Wu, 1993) or frequency domain analysis (Marple, 1987) could be used to process stationary signals (signals with time invariant frequency content), while wavelet (Burrus, *et al*., 1998), or joint time-frequency analyses (Cohen, 1995; Djurdjanovic *et al*., 2002) could be used to describe non-stationary signals (signals with time-varying frequency content). Most real life signals, such as speech, music, machine tool vibration, acoustic emission etc are non-stationary signals, which place strong emphasis on the need for development and utilization of non-stationary signal analysis techniques, such as wavelets, or joint time-frequency analysis.

- **Feature extraction module** extracts features most relevant to describing the product's performance. Those features are extracted from the time domain into which sensory processing module transforms sensory signals, using expert knowledge about the application, or automatic feature selection methods such as roots of the autoregressive time-series model, or time-frequency moments and Singular Value Decomposition.

- **Decision-level sensor fusion** is based on separately assessing and predicting process performance from individual sensor readings and then merging these individual sensor inferences into a multi-sensor assessment and prediction through some averaging technique, or more complex functional transformation.

- **Performance evaluation module** evaluates the overlap between most recently observed signatures and those observed during normal product operation. This overlap is expressed through the so-called Confidence Value (CV), ranging between zero and one, with higher CVs signifying a high overlap, and hence performance closer to normal. In case data associated with some failure mode exist, most recent performance signatures obtained through the signal processing, feature extraction and sensor fusion modules can be matched also against signatures extracted from faulty behavior data. This matching allows the Watchdog Agent™ to recognize and forecast a specific faulty behavior, once a high match with the failure associated signatures is assessed for the current process signatures, or forecasted based on the current and past product's performance. Figure 3 illustrates this signature matching process for performance evaluation. The overlapped areas between the most recent behavior and the nominal behavior as well as the faulty one are continuously transformed into CV over time for evaluating the deviation of the recent behavior from the nominal and faulty ones.

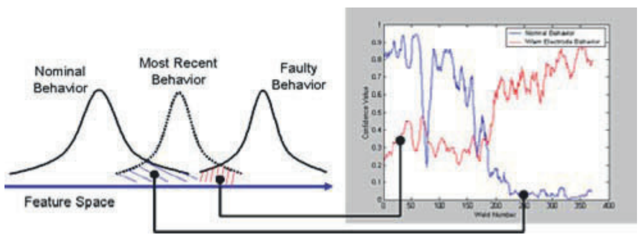


Figure 3. Performance Evaluation using Confidence Value (CV)

Realization of the performance evaluation module depends on the character of the application and extracted performance signatures. If significant application expert knowledge exists, simple but rapid performance assessment based on the feature-level fused multi-sensor information can be made using the relative number of activated cells in the neural network, or using the Logistic Regression approach. For products with open-control architecture, the match between the current and nominal control inputs and performance criteria can also be utilized to assess the product's performance. For more sophisticated applications with intricate and complicated signals and performance signatures, statistical pattern recognition methods, or Feature Map based approach can be employed. Over time, as new failure modes occur, performance signatures related to each specific failure mode can be collected and used to teach the Watchdog Agent to recognize and diagnose that failure mode in the future. Thus, the Watchdog Agent™ is envisioned as an intelligent device that utilizes its experience and human supervisory inputs over time to builds its own expandable and adjustable world model. Even though the performance CV already bares significant prognostic information about the remaining product's useful life, additional prognostic information can be extracted by capturing the dynamics of the product's behavior and utilizing it to extrapolate and predict the product's behavior over time. Currently, Autoregressive Moving-Average (ARMA) modeling and Match Matrix methods are used to forecast the performance behavior.

Furthermore, a system methodology and platform has been developed that enables transformation of product data into more useful formats for remote and telematics applications. The integrated IMS Infotronics platform (Figure 4) has been developed to allow scalability and reconfigurability for many research activities. This platform provides a baseline system for researchers and industry members to develop next-generation intelligent product service systems.

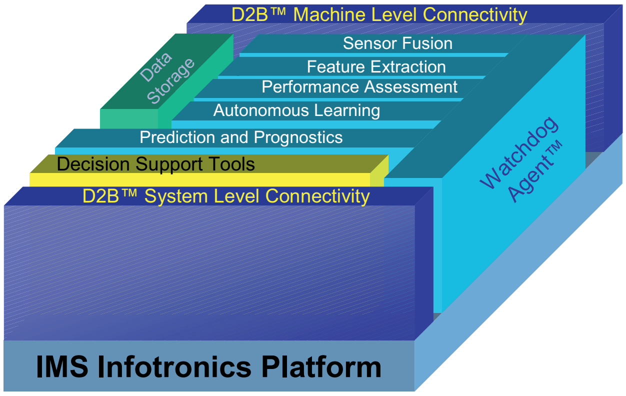


Figure 4. Integrated IMS Infotronics Platform

3.1 Watchdog Agent® Platform Development

3.1.1 Flowchart of developing Watchdog Agent® system

Figure 5 shows the four steps of developing the Watchdog Agent® system for a specific application. Step 1, find out the conditions and needs. Step 2, select the right tool, build and evaluate it. Step 3, Build a prototype system and test it in a real application environment and analyze the results. If the results are not acceptable, then go back to the previous steps. If it's acceptable, go to step 4, build the finalized system and deploy it.

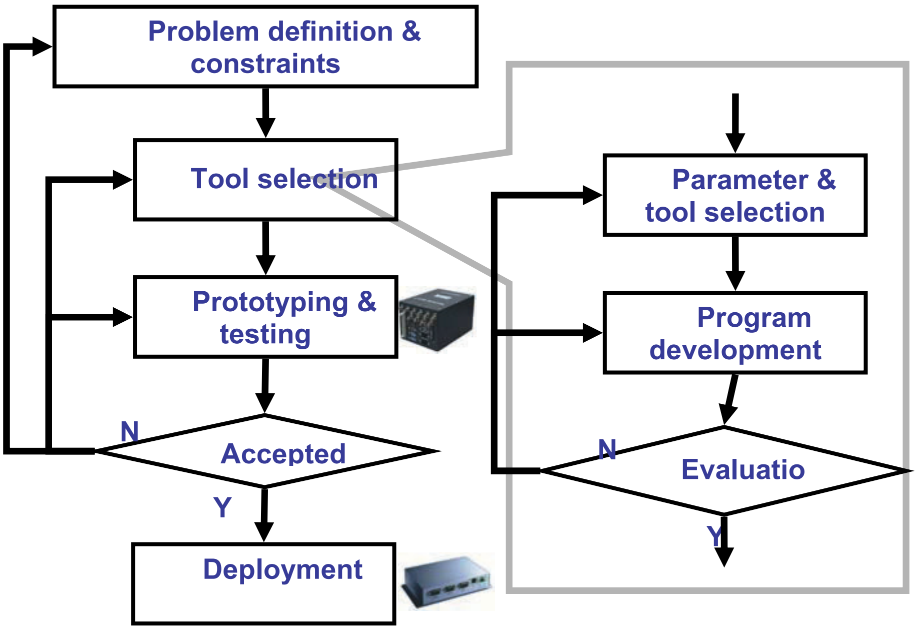


Figure 5. Flowchart of developing Watchdog Agent® system

3.1.2 System Architecture

The Watchdog Agent® system includes embedded computer hardware, an embedded operating system, embedded software and remote client software. The embedded software is comprised of three modules: the Watchdog Agent® tools, the database and the web server. Sensor signals, such as vibration, temperature, pressure, etc, are acquired by the system and processed by the Watchdog Agent tools. Then useful information, such as health information, is saved to the database and published to the internet through an embedded web server. A remote computer may be a remote monitoring workstation, or a maintenance server. The client software we developed runs on a remote computer and gets the useful information via the Internet in order to conduct further analysis.

3.1.3 Hardware Requirements

Hardware requirements depend on the application and the tools we select. For example; an analog input, a digital input, an Ethernet input or a serial input. The tools will be selected based on the computation capacity needs, such as CPU speed, amount of memory and storage. The working environment will decide the storage type, temperature range, etc.

3.1.4 Prototype Hardware Development

Figure 6. Prototype hardware

The prototype hardware is based on PC104 architecture. It has a powerful VIA Eden 400MHz CPU and 128MB memory. It can embed complex algorithms into the hardware. It has sixteen high speed analog input channels to deal with highly dynamic signals. It also has various peripherals that can acquire non-analog sensor signals such as RS-232/485/432, Parallel and USB. PC104 architecture enables the hardware to be easily expanded to multi-board system, which includes multiple CPUs and large amount input channels. The prototype uses a Compact Flash card for storage, so it can be placed on top of machine tools and is suitable for withstanding vibrations in a working environment.

Hardware specification:

- PC104 architecture
- VIA Eden 400 CPU, 128MB RAM
- DAQ features
    - 16-channel, 16 bit 100KHz analog input
    - 4-channel analog output
    - 24 DI/DO, 2 counter/timers
- Dual network interface
- RS-232, USB x 2, VGA
- Spindle-free storage (no HDD)
- Fanless design
- Extended temperature range

3.1.5 Commercially Available Hardware

As an option, we also are evaluating commercially available hardware such as Advantech and National Instruments (NI).

a) Advantech UNO-2160

b) NI-CompactRIO reconfigurable embedded system

Figure 7. UNO-2160 for Watchdog Agent Applications

Figure 8. NI-CompactRIO for Watchdog Agent Applications

3.1.6 Software Development

The software consists of two parts: embedded side software and remote side software, as shown in Figure 9. The embedded side software is the software running on the embedded hardware which consists of the communication module, command analysis module, task module, algorithm module, function module and DAQ module. The communication module is responsible for communicating with the remote side via TCP/IP protocol. The command analysis module is used to analyze different commands coming from the remote side. The task module includes a multi-thread schedule and the management. Algorithm module contains specific watchdog agent tools. The function module has several auxiliary functions such as channel configuration, security configuration, email list and so on. The DAQ module performs A/D conversion using either interrupt or software trigger to get data from different sensors.

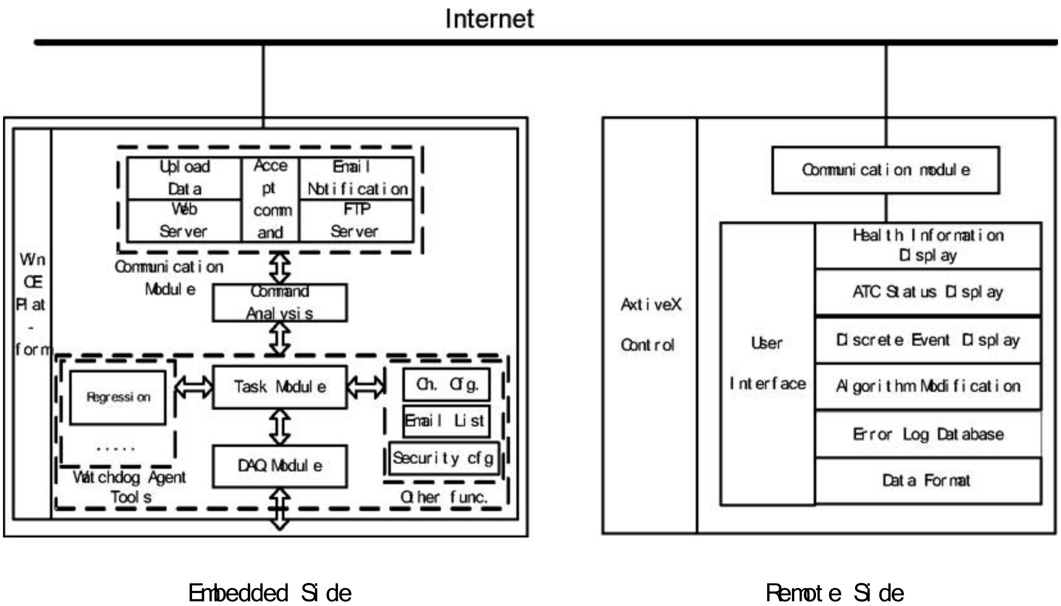


Figure 9 Software structure

The remote side software is the software running on the remote computers. It is implemented by ActiveX control technology and could be used as a component of the Internet Explorer Browser. The remote side software is mainly composed of a communication module and a user interface module. The communication module is used for communicating with the embedded site via TCP/IP protocol. The user interface has a health information display, ATC status display, and a discrete event display, algorithm module, error log database and data format.

4 CASE STUDIES

Several Watchdog Agent™ tools for on-line performance assessment and prediction have already been implemented as standalone applications in a number of industrial and service facilities. Listed below are several examples to illustrate the developed tools.

4.1 Example 1: Prognostics of an AS/RS Materials Handling Systems

A time-frequency based method (Cohen, 1995) has also been implemented for performance assessment of a gearbox in an AS/RS material handling system shown in Figure 11. Four vibration sensor readings have been fused to autonomously and on-line evaluate its performance. The vibration signals were processed into joint time-frequency energy distributions (Cohen, 1995) and a set of time-shift invariant time-frequency moments (Zalubas, *et al*., 1996; Djurdjanovic *et al*. 2000; Tacer and Loughlin, 1996) was extracted. Since those moments asymptotically follow a Gaussian distribution (Zalubas, *et al*., 1996), statistical reasoning was utilized to evaluate the overlap between signatures describing the normal process behavior (used for training) and those describing the most recent process behavior. Figure 12 shows a screenshot of the software application housing this time-frequency based Watchdog Agent used for performance assessment of a material handling system. The CV was generated by fusing multiple signal features for performance assessment.

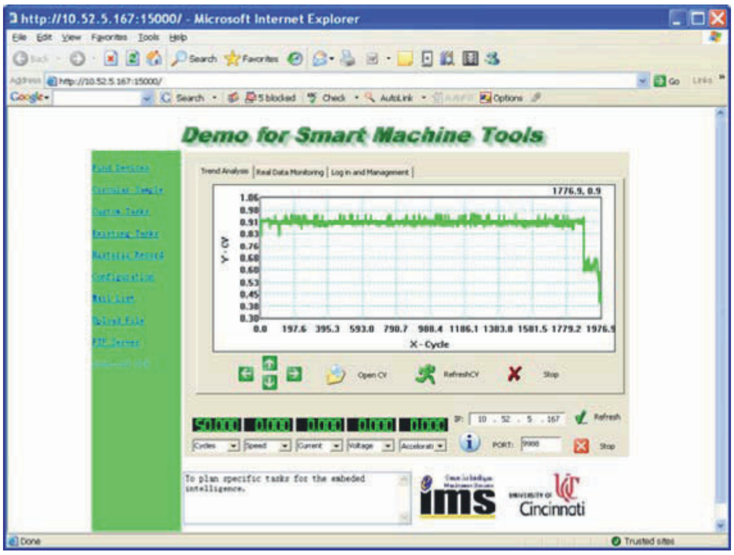


Figure 10. Remote client software

Figure 11. Material Handling System for Mail Staging

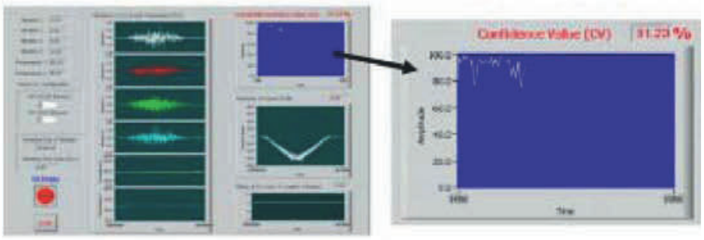


Figure 12. Screenshot of the Time-frequency based Watchdog Agent Assessing Performance of the Materials Handling System Shown in Figure 5

4.2 Example 2: Roller Bearing Prognostics Testbed

Most bearing diagnostics research involves studying the defective bearings recovered from the field, where the bearings exhibit mature faults, or from simulated or "seeded" damage. Experiments using defective bearings have less capability to discover natural defect propagation in the early stages. In order to truly reflect the real defect propagation processes, bearing run-to-failure tests were performed under normal load conditions on a specially designed test rig sponsored by Rexnord Technical Service.

The bearing test rig hosts four test bearings on one shaft. Shaft rotation speed was kept constant at 2000rpm. A radial load of 6000lbs is added to the shaft and bearing by a spring mechanism. A magnetic plug installed in the oil feedback pipe collects debris from the oil as evidence of bearing degradation. The test stops when the accumulated debris adhered to the magnetic plug exceeds a certain level.

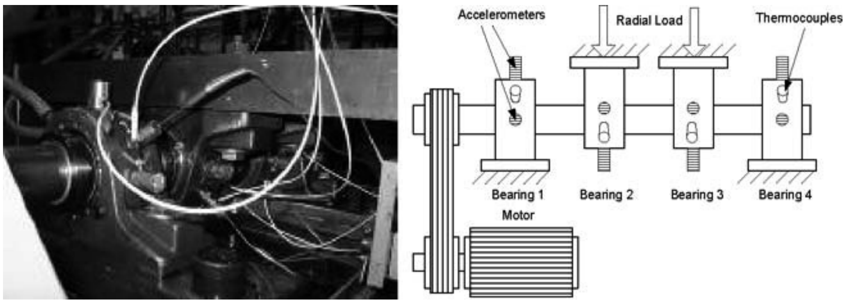


Figure 13. Bearing Test Rig

Four double row bearings were installed on one shaft as shown in Figure 13. High sensitivity accelerometer was installed on each bearing housing. Four thermocouples were attached to the outer race of each bearing to record bearing temperature for monitoring the lubrication purposes. Several sets of tests ending with various failure modes were carried out. The time domain feature shows that most of the bearing fatigue time is consumed during the period of material accumulative damage, while the period of crack propagation and development is relatively short. This means that if the traditional threshold-based condition monitoring approach is used, the response time available for the maintenance crew to respond prior to catastrophic failure after a defect is detected in such bearings is very short. A prognostic approach that can detect the defect at the early stage is demanded so that enough buffer time is available for maintenance and logistical scheduling.

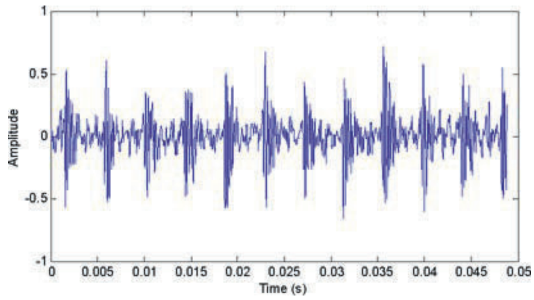
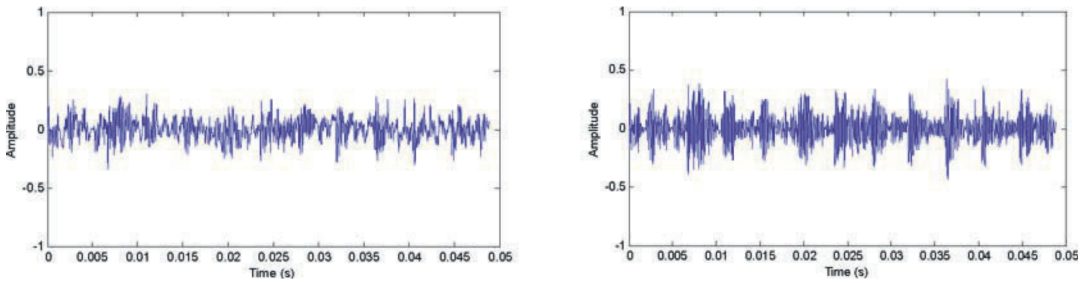


Figure 14. The Vibration Signal Waveform of a Faulty Bearing

Figure 14 presents the vibration waveform collected from bearing 4 at the last stage of the bearing test. The signal exhibits strong impulse periodicity because of the impacts generated by a mature outer race defect. However, when examining the historical data and observing the vibration signal three days before the bearing failed, there is no sign of periodic impulse as shown in Figure 15(a). The periodic impulse feature is totally masked by the noise.



(a) **(b)**

Figure 15. The Vibration Waveform with Early Stage Defect (a) Raw Signal (b) De-noised Signal using the Wavelet Filter

An adaptive wavelet filter is designed to de-noise the raw signal and enhance the degradation detectability. The adaptive wavelet filter is yielded in two steps. At first, the optimal wavelet shape factor is found by the minimal entropy method. Then optimal scale is identified by maximizing the signal periodicity. Applying the designed wavelet filter to the noisy raw signal, the de-noised signal can be obtained as shown in Figure 15(b). The periodic impulse feature is clearly discovered, which is strong evidence of bearing outer race degradation. The wavelet filter-based de-noising method successfully enhanced the signal feature and provided potent proof for prognostic decision-making.

4.3 Example 3: Bearing Risk of Failure and Remaining Useful Life Prediction

An important issue in prognostic technology is to estimate the risk of failure and the remaining useful life of a component given the component age and its past and current operating conditions. In numerous cases, failures are attributed to many correlated degradation processes, which could be reflected by multiple degradation features extracted from sensor signals. These features are the major information regarding the health of the component under monitoring; however, the failure boundary is hard to define using these features. In reality, the same feature vector could be attributed to totally different combinations of the underlying degradation processes and their severity levels. In other words, the failure boundary is gray by monitoring the degradation features. There is only a probabilistic relationship between the component failure and the certain level of degradation features. A typical example can be found during bearing operation. Two bearings of the same type could fail at the different levels of RMS and Kurtosis of vibration signal. To capture the probabilistic relationship between the multiple degradation features and the component failure as well as predict the risk of failure and the remaining useful life, IMS has developed a Proportional Hazards (PH) approach (Liao *et al*. 2005) based on the PH model proposed by Cox (1972). The PH model involving multiple degradation features is given by:

$$\lambda(t;\underline{Z}) = \lambda_0(t)\exp(\underline{\beta}'\underline{Z})$$

where $\lambda(t;\underline{Z})$ is the hazard rate of the component given the current age t and degradation feature vector $\underline{Z}$; $\lambda_0(t)$ is called the baseline hazard rate function; $\underline{\beta}$ is the model parameter vector. This formulation relates the working age and multiple degradation feature to the hazard rate of the component. To estimate the parameters, maximum likelihood approach could be utilized using offline data including the degradation features over time of many components and their failure times. Afterwards, the established model can be used for predicting the risk of failure of the component by plugging in the working age and the degradation features extracted from the on-line sensor signals. In addition, the remaining useful life $L(t_{current})$ given the current working age and the history of degradation features can be estimated by:

$$L(t_{current}) \approx \int_{t_{current}}^{\infty} \exp\left(-\int_{t_{current}}^{\tau} \lambda(v;\underline{\hat{z}}(v))\, dv\right) d\tau$$

where $\underline{\hat{z}}(v)$ is the predicted feature vector.

Consider the vibration data obtained from the test rig in Example 2. To facilitate on-line implementation, root-mean-square (RMS) and Kurtosis are calculated and used as degradation feature. Figure 10 shows the predicted hazard rate over time based on these degradation features. This quantity can be utilized to trigger maintenance when the risk level crosses a predetermined threshold level. Table 1 provides the remaining useful life predictions given the current bearing age and the feature observations. The predictions are in accordance with the actual life of the studied bearing (32.5278 days) with minor prediction errors when the degradation progresses.

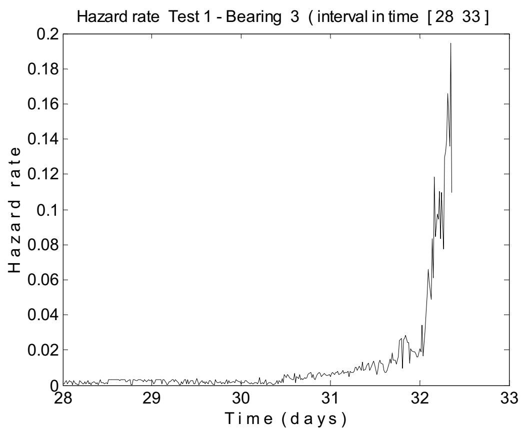


Figure 16. Hazard Rate Prediction of Bearing 3 in Test 1

Table 1. Estimates of Expected Remaining Useful Life – Test 1, Bearing 3

Time *	26	29	31
Estimated expected remaining useful life*	3.5549	3.3965	1.5295
True remaining useful life *	6.5278	3.5278	1.5278
Error *	2.9729	0.1313	0.0017

* in days

5 CONCLUSIONS AND FUTURE RESEARCH

Watchdog Agent™ is a tool for multi-sensor assessment and prediction of machine or process performance. This tool can be utilized to realize predictive Condition-Based Maintenance as well as identification of components with significant remaining useful life. A wide variety of Watchdog Agent™ realizations have been devised to address applications of different nature and different levels of complexity and criticality. Innovative sensory processing and autonomous feature extraction methods are developed to facilitate the plug-and-play approach in which Watchdog Agent™ can be set and ran without any need for expert knowledge or intervention. Watchdog Agents™ have been successfully implemented in industrial testbeds.

6 REFERENCES

[1] Burrus C, Gopinath R, Haitao G. Introduction to wavelets and wavelet transforms - a primer. NJ: Prentice Hall; 1998.

[2] Casoetto N, Djurdjanovic D, Mayor R, Lee J, Ni J. Multisensor process performance assessment through the use of autoregressive modeling and feature maps. Trans. of SME/NAMRI 2003; 31:483-490.

[3] Cohen L. Time-frequency analysis. NJ: Prentice Hall; 1995.

[4] Cox D. Regression models and life tables (with discussion), Journal of the Royal Statistical Society, Series B 1972; 34:187-220.

[5] Djurdjanovic D, *et al*. Computerized classification of temporomandibular joint sounds. IEEE Transactions on Biomedical Engineering 2000; 47:977-984.

[6] Djurdjanovic D, Ni J, Lee J. Time-frequency based sensor fusion in the assessment and monitoring of machine performance degradation. Proceedings of 2002 ASME Int. Mechanical Eng. Congress and Exposition, paper number IMECE2002-32032.

[7] Goodenow T, Hardman W, Karchnak M. Acoustic emissions in broadband vibration as an indicator of bearing stress. Proceedings of IEEE Aerospace Conference, 2000; 6: 95-122.

[8] Hansen R, Hall D, Kurtz S. New approach to the challenge of machinery prognostics. Proceedings of the International Gas Turbine and Aeroengine Congress and Exposition, American Society of Mechanical Engineers, June 13-16 1994; p.1-8.

[9] Hardman W, Hess A, Sheaffer. A helicopter powertrain diagnostics and prognostics demonstration. Proceedings of IEEE Aerospace Conference, 2000, 6:355-366.

[10] IMS, NSF I/UCRC Center for Intelligent Maintenance Systems, www.imscenter.net; 2004.

[11] Kemerait R. New cepstral approach for prognostic maintenance of cyclic machinery. IEEE SOUTHEASTCON, 1987; p. 256-262.

[12] Kleinbaum D. Logistic regression. New York: Springer-Verlag; 1994.

[13] Lee J. Machine performance monitoring and proactive maintenance in computer-integrated manufacturing: review and perspective. International Journal of Computer Integrated Manufacturing 1995; 8:370-380.

[14] Lee J. Measurement of machine performance degradation using a neural network model. Computers in Industry 1996; 30:193-209.

[15] Lee J, Ni J. Infotronics agent for tether-free prognostics. Proceeding of AAAI Spring Symposium on Information Refinement and Revision for Decision Making: Modeling for Diagnostics, Prognostics, and Prediction, Stanford Univ., Palo Alto, CA, March 25-27, 2002.

[16] Liao H, Lin D, Qiu H, Banjevic D, Jardine A, Lee J. A predictive tool for remaining useful life estimation of rotating machinery components. ASME International 20th Biennial Conference on Mechanical Vibration and Noise, Long Beach, CA, 24-28 September, 2005.

[17] Marple S. Digital spectral analysis. NJ: Prentice Hall; 1987.

[18] Pandit S, Wu S-M. Time series and system analysis with application. FL: Krieger Publishing Co.; 1993.

[19] Radjou N. The collaborative product life-cycle. Forrester Research, May 2002.

[20] Reichard K, Van Dyke M, Maynard K. Application of sensor fusion and signal classification techniques in a distributed machinery condition monitoring system. Proceedings of SPIE - The International Society for Optical Engineering, 2000; 4051:329-336.

[21] Roemer M, Kacprzynski G, Orsagh R. Assessment of data and knowledge fusion strategies for prognostics and health management. Proceedings of IEEE Aerospace Conference, 2001; 6:62979-62988.

[22] Seliger G, *et al*. Innovative processes and tools for disassembly. Annals of CIRP 2002; 51:37-41.

[23] Su L, Nolan M, DeMare G, Carey D. Prognostics framework [for weapon systems health monitoring]. Proceedings of IEEE Systems Readiness Technology Conference, IEEE AUTOTESTCON '99, 30 August-2 September 1999; p. 661-672.

[24] Tacer B, Loughlin P. Time-frequency based classification. SPIE Proceedings, 1996; 42:2697-2705.

[25] Wilson B, *et al*. Development of a modular in-situ oil analysis prognostic system. International Society of Logistics (SOLE) 1999 Symposium, Nevada, Las Vegas, 30 August –2 September, 1999.

[26] Yamada A, Takata S. Reliability improvement of industrial robots by optimizing operation plans based on deterioration evaluation. Annals of the CIRP 2002; 51:319-322.

[27] Yen G, Lin K. Wavelet packet feature extraction for vibration monitoring. IEEE Trans. on Industrial Electronics 2000; 47:650-667.

[28] Zalubas E, *et al*. Shift and scale invariant detection. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 1996; 5:3637-3640.

ASSET MANAGEMENT – AN INSURANCE PERSPECTIVE

Ian Barnard

Senior Risk Management Engineer - American International Group

Abstract: The global insurance industry is vitally interested in the Asset Management systems that are implemented at industrial sites. While these systems are usually implemented to protect and manage plant and equipment, often the impact of these systems on the corporate financial model are overlooked. The type and effectiveness of the asset management system can affect the companies viability in terms of insurance cover and hence capital raising and risk profiling. This paper discusses the operation of the insurance market and the impact that asset management systems have on both the insurer and the insured.

Key Words: insurance, plant, asset management systems

1 BACKGROUND

The insurance industry is a global business centred in New York, London, Zurich and the Bahamas. Worldwide, insurance premiums in 2004 exceeded $3 trillion. In the United States, which accounts for about 34 percent of the world's insurance business, premiums total in excess of $1 trillion. The insurance industry is generally viewed as three distinct segments, property / casualty, life and health. The property / casualty part of the industry provides insurance for cars, homes and businesses. The term "casualty" dates back to the time before the 1950s when property/casualty insurers were two distinct kinds of insurance companies, with casualty or liability insurers covering losses that resulted from casualties and property insurers covering damage to or loss of property.

Property/casualty insurance can be broken down into two major categories: commercial lines or types of insurance and personal lines. Personal lines, as the term suggests, includes coverage for individuals –- car and homeowners insurance. Commercial lines, which accounts for more than half of U.S. property / casualty insurance industry premiums, includes the many kinds of insurance products designed for businesses.

Commercial insurance performs a critical role in the world economy. Without it, the economy could not function. Insurers essentially protect the economic system from failure by assuming the risks inherent in the production of goods and services. This transfer of risk frees insured companies from the potentially paralysing fear that an accident or mistake could cause large losses or even financial ruin. As a consequence, the Insurance industry's interest in asset management systems is very real and very focused.

Managing risk requires an in depth knowledge of the systems and processes utilised by a large variety of businesses. The North American Industrial Classification System identifies some 1,170 different industries according to the processes used to produce goods or services. In the United States, there are some seven million business establishments, each of which employs one or more of these processes and each of which buys some kind of insurance.

AIG and all comparable insurance companies collect plant and system root cause failure analysis data, including type faults which may occur in established and new plant and equipment. This data is used as reference material for this paper.

2 THE WAY THE INDUSTRY WORKS

2.1 Capacity

Each insurance company has a finite level of risk (financial reserves) that can be allocated to a given project or industrial complex. This risk is termed Capacity, and the magnitude of the capacity varies with the size of the company in terms of its financial reserves, and the risk profile that the insurer currently has on its books. Companies vary their capacity in proportion to the level of risk already on their books in a given industry sector or geographical region. In order to secure the full insurance cover required by large businesses, insurance risk is spread over a number of insurance companies.

Every policy has a lead insurer, who will negotiate terms and conditions with the insured or their agents (Brokers). The lead insurer takes the first and highest probability risks for the site, and allocates a certain capacity to that risk, at a certain premium.

Other companies are then invited to participate in the risk, and their involvement will be initiated if / when the losses for an event exceed a predetermined level. In this way, the insurance book is filled for a give client.

The globalisation of the insurance industry can be grasped by examining a typical (simplified) portfolio of insurance providers for a major industrial asset.

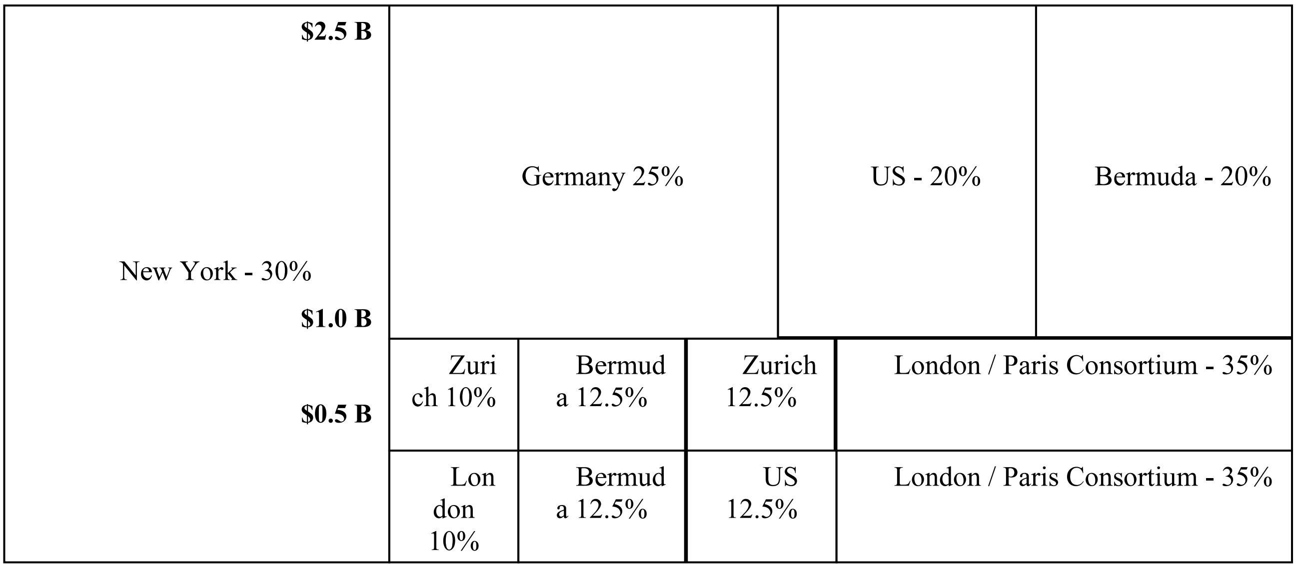


Figure 1: Typical Insurance Placement Map

The diagram depicts the levels of insurance cover and the companies that offer that cover. The Y axis depicts the total level of cover for the industrial plant, which in this case is $2.5B. The X axis is complied from companies who agree to cover certain percentages of the risk up to defined limits. In this example, a consortium of companies will provide cover to this risk up to the first $500 M. A different consortium of companies will provide cover for the second $500M, taking the combined cover to $1B. Likewise, the final $1.5B cover will be provided by a third and different set of companies. Each company will charge a premium based on their level of exposure, and where they are placed in that cover, be it the first, second, third or more tier. Obviously, the companies in the first tier of cover are exposed to the greatest risk, and so the price for this tier is higher than for subsequent tiers. The first New York company, who is providing cover over the whole risk, is the lead insurer. However, this company will also offset its exposure by utilising the services of one or more Reinsurance companies, effectively “buying” more capacity and divesting some of its risk to third parties. Other members of each tier may also utilise reinsurance.

The point of this diagram is that there are a large number of decision makers involved in determining not only the cost of insurance to a particular project, but also deciding on their involvement or otherwise in the risk. This means that, for a large asset, there are many people and organisations that need to be convinced of the soundness of the insured’s Asset Management strategies. Unless all parties are satisfied with the risk, the facility will not be insured.

2.2 Typical Plant Policy Structure

Insurance for a large industrial plant is generally structured so that relatively minor losses are born solely by the company which owns the plant. The typical structure for a power generation facility, for example, is that only losses above $5M are claimable, and interruptions to business of less than 45 days are not insurable. This fact alone has a significant impact on the way insurance companies view asset management systems. In general, from a loss viewpoint, asset management systems which protect small and / or low impact systems are not on the Insurance radar.

Premiums will generally run at between 1% and 2% of the insurable loss in first world politically stable economies. For a typical large power generation facility with a capacity of 2000 MW, the premium will be somewhere between $2M and $4M.

2.3 Loss Probabilities

Loss probabilities for a risk are calculated based on the likelihood of certain events. Following are the standard loss definitions with examples for a power station.

NLE: Normal Loss Expectancy

This type of loss is based on a normal event that has a reasonably high likelihood of occurring in a plant. While the event will affect production, the scenario is reasonably common and in general is not covered by insurance given that the policy deductible (excess) will exceed the cost of rectification.

Plant	Boiler	Turbine
Event	Mild Furnace Explosion	Steam turbine incident
Property Damage	$250,000	$2.7M
Business Interruption	5 days	30 days
BI (% Prod. Effected)	25%	25%
$BI	$250,000	$1.5M
TOTAL ($)	$500,000	$4.2M

PML: Probably Maximum Loss

This type of event is not expected to occur under normal condition, but experience shows that it has a finite probability of eventuating over the life time of the asset. The losses associated with this type of event are large, and rhe values are calculated assuming that all installed protection and asset management systems function as designed to limit the impact of the event.

EVENT	Low boiler water	Fire resulting in loss of rotor
Property Damage	$25M	$36M
Business Interruption	180 days	365 days
BI (% Prod. Effected)	25%	25%
$BI	$9M	$18M
TOTAL ($)	$34M	$54

EML – Expected Maximum Loss

This is the loss that can be expected to occur under the worst possible conditions for a plant and where the protection systems designed to minimise the loss fail completely. The loss is limited only by passive protection such as plant spacing. This is the largest of the possible losses, and the values associated with this type of loss are greatly affected by inherent risk mitigation such as building and processes separation etc.

EVENT	Boiler explosion	Catastrophic fire
Property Damage	$160M	$85M
Business Interruption	30 months	30 Months
BI (% Prod. Effected)	25%	25%
BI	$45M	$45M
TOTAL ($)	$205M	$130M

Despite the significant level of research and technical analysis that is applied to determine the risk for a given project, often the decision by a particular insurance company to be involved (and effectively investing in) a project is not fully quantifiable. Insurance professionals are in the main financial specialists, and are not engineers. Traditionally, they were concerned with questions of political and market stability, natural perils such as flood and earthquake, facility fire protection and a host of other risks to a project or site. While the actual numbers generated statistically to evaluate a risk profile are important to a decision, there is a certain amount of qualitative analysis that occurs, and this is best described by intangibles such as the experience of an underwriter in a particular risk field. Not only do the numbers for a risk have to be right, but the “feel” of the risk has to be right.

In recent years, the profile of the commercial and industrial insurance industry has developed to include risks far removed from the traditional insurance cover.

3 INSURANCE TYPES

There are almost as many types of insurance as there are insurance companies. By type of insurance we mean the types of risks covered. As stated previously, the types of insurance can be divided into Personal and Commercial. For our purposes, we will concentrate on the insurance that is concerned with commercial business assets.

3.1 Property Insurance

Traditionally, insurance cover was primarily related to the destruction of a facility or assets by either fire or natural disaster. This type of insurance is known in the industry as Property cover, and as the name suggests it relates principally to the replacement value of real property. Impacts here include political, geological, fire, explosion and other major issues that can destroy of severely damage the physical plant as a whole. The insurance industry is well versed in analysing and assessing the risks related to these issues, and various statistical tools are used to quantify the risks from these aspects of an operation.

3.2 Boiler and Machinery Insurance (BM)

This type of insurance, despite the name, relates to the risk of plant being destroyed or rendered inoperable from any source. In effect, it is Reliability Insurance for the plant, and takes into account aspects such as premature failure of major systems, the destruction of plant from fire, design defects and a myriad of other causes. It is interesting to note that stupidity is insurable, laziness is not. By this we mean that damage due to honest mistakes by operators or maintainers will be covered by the policy. Damage resulting from poor maintenance and inspection practices, plant over load, operation outside design parameters and a number of other inferior practices are not covered by these policies.

3.3 Business Interruption Insurance (BI)

BI Insurance covers the loss of profit risk to a project or site from an interruption to the supply chain. It is possible for this type of insurance risk to exceed the risk from all other sources. For example, the destruction of a compressor in a petrochemical plant from a catastrophic rotor failure may require a complete compressor replacement costing approximately $1M. However, if the compressor is a purpose built machine, with a replacement time of say 6 months, the loss of profit to the organisation may exceed the cost of replacement by orders of magnitude. Business interruption insurance has a critical interest in the quality of the asset management systems applied to a process.

BI insurance does not include damage related to wear and tear, or plant and equipment that is not maintained and inspected to industry standard. This aspect has connotations for asset management systems designers, as the term “industry standard” has no global reference, and varies from industry to industry and country to country.

4 RISK RATINGS

Asset management system design and implementation has a crucial impact on both Boiler and Machinery and Business Interruption insurance. Insurance companies assess the standard of asset management applied to a particular site, and decide not only on the cost of such cover, but on the provision of such cover based on engineering reports compiled on the project.

Engineering reports are produced by qualified and experienced engineers who visit and review a site or project, and inspect the systems in place at the site to minimise risk. The systems review includes human and technical systems, and looks at both the design of systems and the level of effective implementation of the system. At AIG and in other major insurance companies, plants are categorised and rated as Excellent, Good, Fair and Poor risks based on a quantitative and qualitative assessment of the risk management systems implemented at the site. The headings used to evaluate the risks for an engineering study are:

Location, Plant Layout, Fuel Storage and Handling, <u>Process Control</u>, <u>Reliability</u>, <u>Maintenance</u>, <u>Inspection</u>, Operation, Fire Protection, <u>Risk Management</u>. The underlined topics are critically dependant on asset management technologies.

As an example, below is a sample of grading standards for the Process Control element in a power station as determined by AIG. Grade 0 is excellent, Grades 1 and 2 are Good, Grade 3 is Fair, Grade 4 is Poor.

Process control	**Grading**
Control rooms & systems:	Grade 0: Process control DCS separate from critical ESD and trip systems, hard wired or Triple Modular Redundancy system, input voting system. Grade 1: As above, but DCS and ESD are linked and dependent. Grade 2: Manually actuated automatic shut down of process unit, with satisfactory level of auto shut down of unit packages such as rotating equipment. Grade 3: Mix of old and new systems but not integrated. Unmanned stations. Grade 4: Local controls only
Boiler	Grade 0: fully automatic combustion controls on all burners. Alarms interlocks. Drum level trips, 2 out of 3 voting system, separate level indication in control room (CCTV). Double isolation valve on gas or oil feed. Grade 1: some manual controls on non critical processes Grade 2: strict purge sequence, drum level trip, burner management system Grade 3: no burner management system Grade 4: no burner management system, no combustion control, manual light up
Turbine	Grade 0 : Double overspeed protection system, remote condition monitoring, on-line protection facility. Fully automated governor control with stress control (ST). On line gas analysis for GT Grade 1 : On line vibration measurement with trips and trending Grade 2 : On line vibration and eccentricity with trips. Single overspeed protection system. Grade 3 : On line vibration and eccentricity without trips. Grade 4 : No condition monitoring system (ST).
Generator	Grade 0: generator core monitoring, on line partial discharge monitoring, phase and balance protection, reverse power protection Grade 1: As above but no online PD Grade 2: See best practice page 23 (minimum practice). Grade 3:

Figure 1: Excellence Grades for Power Station Process Control Element

From these ratings, an overall risk management profile and rating is developed.

Facilities placed in the Excellent category have no trouble obtaining insurance cover, and will have the lowest premiums and the least stringent policy conditions. These companies will have significant bargaining power in the insurance market and may even be able to dictate their own terms to the global insurance market.

Industrial sites placed in the Good category will obtain insurance on terms that reflect the lower rating. Premiums are likely to be higher and insurance companies may impose restrictions and riders that limit the level of cover. In many cases, additional risk management and asset protection systems will be specified by insurance companies before cover is applicable. In some cases, these additional systems can cost millions of dollars.

Facilities in the Fair and Poor categories will find insurance difficult to obtain, and this will effect the viability of the operation as a whole. Banks will not finance companies who cannot obtain insurance. If insurance companies are at all interested in being involved in the facility, a whole range of management systems may be mandated prior to cover being negotiated, and the premiums for the insurance will reflect the poor risk profile.

5 PLANT FAILURES

From an asset management perspective, it is necessary to understand the root cause of plant failure in order establish effective condition monitoring and asset management strategies. AIG and other organisations have conducted considerable research into the reasons for plant failure. Not surprisingly, many plant failure studies into mechanical and electrical equipment failures reveal similar system failure root causes, which are indicated in the following graphics.

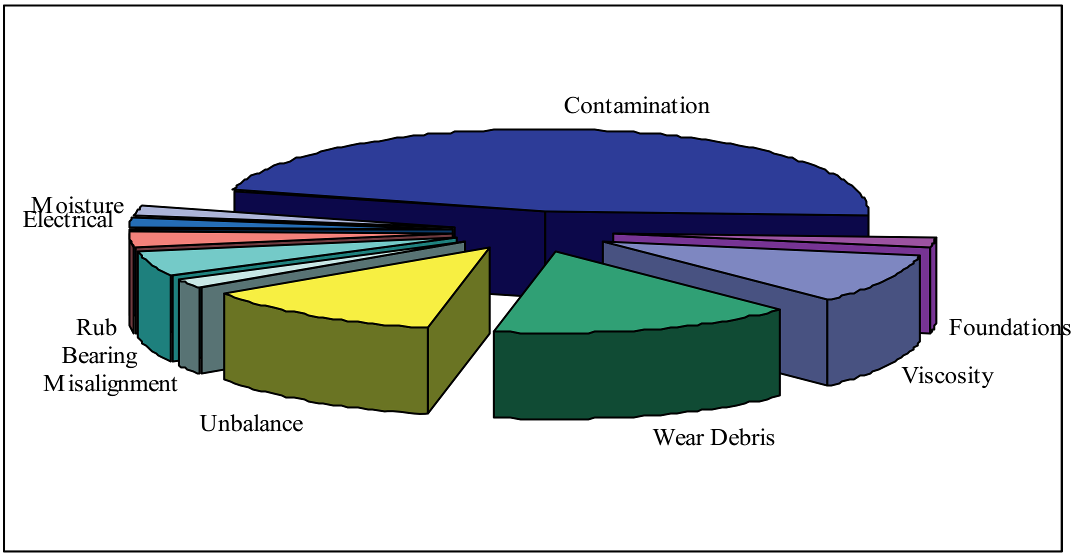


Figure 2: Primary Causes of Mechanical Equipment Failure

This data has been taken from a number of industries of varying ages and capacity factors. In general, the data indicates that human intervention in the machine in the form of lubrication quality and contamination control is the primary factor in the longevity of mechanical plant. The data also indicates that lubrication analysis is the prime control method of determining mechanical plant condition, particularly in rotating mechanical plant.

The following graphic depicts the systems within electrical plant that have been the principle failure cause for that plant. It is interesting to note that mechanical systems within electrical plant are the prime failure areas. The obvious connotation from this data is that the complete suite of both mechanical and electrical testing and condition determination regimes must be applied to all plant types. The complimentary nature of the condition assessment techniques requires that the systems used to record and report on plant condition need to be fully integrated, and contain data, analysis and prognosis from disparate CM tools to give a complete description of the plant condition.

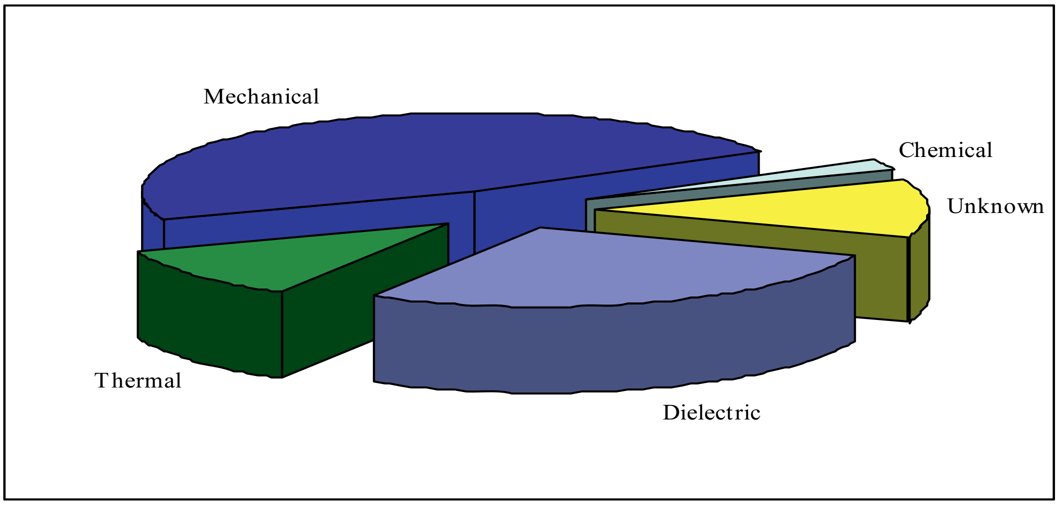


Figure 3: Primary Causes of Electrical Plant Failure

With these failure modes in mind, to obtain an excellent rating from insurers the asset management and condition monitoring technologies that are required for a utility include:

Mechanical:
- Online vibration spectral analysis of main plant including turbines, generators, feed pumps.
- Online vibration measurement and trips on auxiliary plant.
- Spectro chemical and Microscopic wear debris analysis of all main lube oil systems.
- Infra red thermography of all main plant including turbine casings, boilers, auxiliary plant bearings.
- Eddy current testing of condensers
- Time of flight ultrasonics (TOFD)
- Eddy current & remote field testing (EC & RFT)
- Remote visual inspection (RVI)
- Replication
- Bore oxide ultrasonics
- Mini creep testing
- X Ray analysis

Electrical:
- Dissolved gas analysis (DGA)
- Dissolved moisture
- Oil quality (acid, interfacial tension, dielectric strength, dielectric dissipation factor)
- Particle count
- Dissolved metals and sulphur
- Furans
- Dielectric loss angle
- Insulation resistance
- Dielectric dissipation factor
- Insulation recovery voltage
- Partial discharge
- Time reflectometry
- Radio frequency emission
- Transfer function

Risk Assessment Tools
- HAZOPS - Hazard & Operability Studies
- QRA - Quantitative Risk Assessment
- WRAC - Workplace Risk Assessment and Control
- RCM-RA - Reliability Centred Maintenance Risk Assessment

It should be carefully noted however that the existence of these systems at a site is NOT the determining factor for a risk management engineer. The critical aspect is how these systems are utilised in the asset management planning cycle. All of the above systems require analysis and prognosis development by humans, and of crucial importance to the insurance industry are the qualifications and experience of the people conducting the testing and performing the analysis.

6 IMPACTS FOR ASSET MANAGEMENT SYSTEM DESIGN

The insurance industry is both a friend and foe to the designers and researchers involved with asset management systems.

Firstly, the scale of the asset management systems from an insurance perspective is company and / or plant wide. The levels of insurance deductible (excess) imposed means that companies effectively self insure for the first $5M, and cover their own business profit losses for the first 45 days . Asset management systems which protect plant with an impact less than these figure are generally not important to the insurance industry. It should be noted however that risks which are placed in the Excellent and Good ranges by the insurance Risk Consultant as a result of superior asset management systems can often effectively negotiate the deductible and BI exclusion period..

Secondly, to obtain Business Interruption insurance for example, insurance policies mandate that an insured company follows a number of guidelines, including adherence to the equipment manufacturers operations and maintenance schedules and compliance with industry normal practice for the asset. This has the effect of greatly limiting the scope for innovation to be applied in asset management techniques and methodologies from an insurance premium reduction perspective.

On the other hand, the insurance companies also mandate that traditional asset management and condition monitoring methods be applied. For example, the application of thermography, oil and vibration analysis, dissolved gas analysis, partial discharge testing, the full suite of non destructive metallurgical techniques and a host of others are all required by business interruption insurance providers to protect the assets.

The message here is that's its is necessary to utilise proven CM techniques to manage assets, but don't rely on anything too innovative and unproven.

Risk management engineers are usually assigned to a client for the duration of the insurance relationship. These people are the front line of the technical interface between the insurance portfolio developed for the client and the asset management processes implemented by the client. They are in a position to suggest additional asset management technologies that would benefit the clients business and reduce the risk of breakdown and business interruption.

The point here is that a direct link between emerging asset management technologies and the potential market for those technologies exists through the insurance chain. This link has been largely unutilised by the asset management community. If asset management researchers and technology developers wish to get their work acknowledged and into the market, the insurance industry in general and the risk management consultant in particular can be a valuable ally, provided that the technology has demonstrable advantages to both the insured and insurer.

While it is not possible to quantify the effect that the installation of a particular technology at a site will have on an insurance premiums, it is possible to move a risk from a Fair or Poor risk rating to a Good or Excellent rating, with a consequential improvement in the companies insurance profile. In some cases, the implementation of additional risk management systems will determine the viability of the whole project.

7 HUMAN ELEMENT OF ASSET MANAGEMENT

All those who work in the Asset Management field know of examples where effective asset management technologies have been implemented only to fail in the field due to a lack of integration with existing systems and processes. The assessment by insurers of asset management systems does not rely on the presence of installed technology, or the levels of capital investment by a company in asset management technologies. The assessment of the asset management standard of a plant is based solely on how the technologies are utilised to reduce the risk of plant failure.

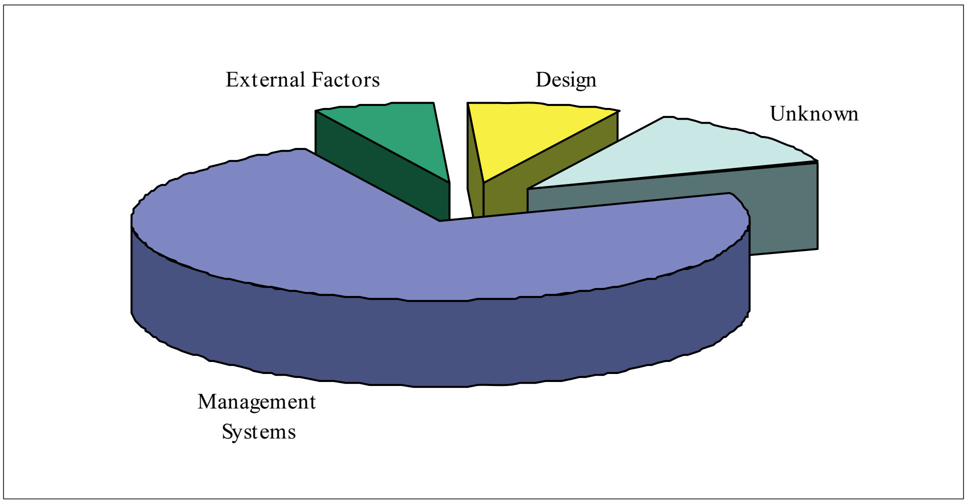


Figure 4: Upper Level Systems Failure Analysis

The following graph depicts the major causes (from a systems failure viewpoint) of insurance losses experienced in the US in the power industry in 2005. While all these losses can be analysed to identify the aspects of the mechanical and electrical systems that failed, the losses have been further analysed to determine if a failure in asset management systems contributed to the level of loss incurred.

In this graph, Management Systems denotes all types of risk management systems, including HAZOP studies, plant inspections, condition monitoring systems and routine plant inspections. External factors include incidents that occurred outside the plant over which the plant management had no control.

The data indicates that approximately 75% of losses resulted from a failure to implement or sustain targeted and robust plant asset management systems.

The data was further analysed to ascertain the most appropriate loss prevention strategy that could have been implemented to reduce that particular risk of failure in the plant.

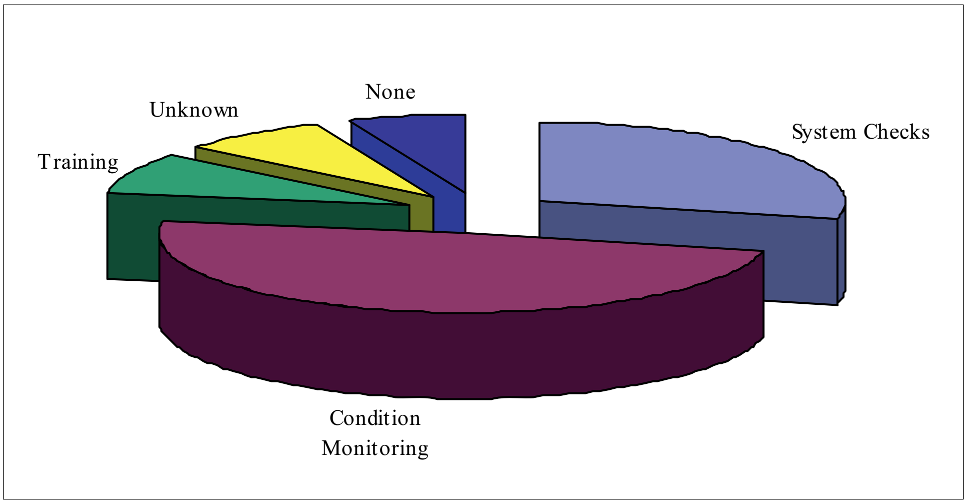


Figure 5 : Possible Loss Prevention Strategies

This data indicates that approximately 50% of the loss values could have been avoided or reduced if the appropriate plant condition monitoring systems had been implemented. A further 30% of the losses could have been favourably impacted by the application of simple plant inspection regimes. It is interesting to note that 6% of the losses incurred could not have been mitigated using technologies available today, and in a further 8% the system failure at the root of the loss is unknown.

This data, plus large amounts of anecdotal evidence, suggests that it is not the development of additional asset management technologies that will have the greatest impact the levels of losses in this industry. Rather, it is the development of systems which are fully integrated into the plant operations and maintenance process that will reduce the levels of plant downtime. This is not to say that the development of new techniques in plant condition monitoring are not important or required. Many companies and industries are in dire need of specialist testing techniques to enable product improvement and to increase reliability. However, Insurance companies are primarily interested in systems that deliver real reliability and availability returns on a plant wide basis.

8 CONCLUSIONS

The information presented above includes some salient points for Asset Management system designers:

- Rightly or wrongly, the insurance industry has considerable influence on the types of asset management systems implemented in a commercial enterprise. This fact can be used to advantage by asset management system developers.
- Insurance underwriters rely almost completely on the information supplied by relatively few risk management engineers to determine the adequacy of the asset management systems installed at a facility.
- In the main, the insurance industry does not financially support the development of asset management technologies. The industry is principally geared towards the implementation of those technologies and systems which are proven to improve the reliability of plant and equipment.

- While it seems incongruous with the previous point, the insurance industry is vitally interested in technologies which reduce the risk of business interruption to plant and processes.

- Insurance company Risk Management Engineers may be a valuable aid in implementing innovative asset management and plant condition monitoring technologies.

9 REFERENCES

[1] Insurance Information Institute of America : Yearbook 2005

[2] AIG Risk Management Handbook, 2003

[3] AIG Best Practice – Utilities, 2004

[4] AIG Engineering Manual

[5] Pacific Power Condition Monitoring Handbook – 1999

[6] AIU Field Engineering Handbook

[7] Munich Re Technical Insurance References – 2005

[8] AIG 2005 Major Plant Losses – Proprietary AIG data

[9] Evans B - From Planned Maintenance to Condition Based Maintenance - ICM Maintenance Management Summit – Sept 99

[10] Bissett W - Working Smarter Through Condition Monitoring – Pacific Power International - 2002

CHALLENGES IN ACHIEVING OPTIMAL ASSET PERFORMANCE BASED ON TOTAL COST OF OWNERSHIP

Gautam Sardar [a], Ramachandran N. [b], and Ravi Gopinath [c]

. [a] *Dr. Gautam Sardar: Head, Research and Development, Engineering and Industrial Services, Tata Consultancy Services Limited, INDIA, Email: gautam.sardar@tcs.com*

[b] *Dr. N. Ramachandran: Head, Enterprise Asset Management Practice, Engineering and Industrial Services, Tata Consultancy Services Limited, INDIA, Email: ramachandran.n@tcs.com.*

[c] *Dr. Ravi Gopinath (Keynote Speaker): Vice President, Engineering and Industrial Services, Tata Consultancy Services Limited, INDIA, Email: ravi.gopinath@tcs.com.*

Abstract: Ensuring optimal performance of assets in capital-intensive industry sectors is crucial towards sustained growth in terms of returns on investment (ROI). The primary challenge is to achieve optimal utilization of the assets over their productive lifecycle and identify the need and timing for replacement and/or enhancement. Such decisions are not trivial and can affect the ROI considerably. Operational and planning decisions related to the assets need to be based on the concept of Total Cost of Ownership (TCO) if they have to be economically optimal. The technical challenges posed are in terms of selection of a mix of techniques and methods for extracting useful information/knowledge from raw operational data, its use to construct predictive models for asset performance, and use of the models to achieve continuous optimization of asset performance with TCO consideration. It requires a model based approach to achieve asset performance over multiple lifecycles. Technology enablers are also a crucial element in achieving the maturity of asset management processes and practices in the organization. The paper discusses an EAM maturity model and its technological requirements that would help organizations to achieve sustained asset performance based on TCO.

Key Words: Enterprise Asset Management (EAM), Total Cost of Ownership (TCO), Optimization, Asset Performance, Model Based Approach, Process Maturity

1 INTRODUCTION

Enterprise Asset Management (EAM) is traditionally defined as the organized and systematic tracking of an organization's physical assets such as its plant, equipment and facilities. The machinery used to manufacture the products has to be reliable, the electricity must flow, the computers must remain on line, and the delivery trucks must be operational. If just one of these assets breaks down unexpectedly, the operations may be interrupted and this can impact organization’s bottom line. In its earlier avatar of Maintenance Management, the thought revolved primarily around Maintenance, Repair, and Operations (MRO). This was limited to fixing any sort of equipment should it get out of order or broken (repair) as well as performing the routine actions which keep it in working order (maintenance) or prevent trouble from arising (preventive maintenance). Preventive Maintenance focused on the activities required to continuously monitor health of the equipment and plan timely maintenance ensuring its availability as per the organizational requirement. The approach and underlying thought process sufficed to support the typical organizational thought process of doing business with high levels of profitability and operational efficiency. The MRO philosophy is sufficient from the perspective of driving operational efficiencies and deriving appropriate benefits out of that to optimize the profitability on ongoing basis.

One of the reports by the National Institute of Standards and Technology (NIST GCR 04-867) estimates that the U. S. Capital Facilities Industry loses over $15.8 billion annually due to inefficient exchange and management of design and engineering information. This is a glimpse of the size of benefits that can be achieved through systematic asset management [3].

This paper aims a discussion on holistic approach to Enterprise Asset Management from all the important dimensions of incorporated into the asset’s product lifecycle. The thrust of this discussion is to understand the effect of these dimensions on EAM and handling of various trade-offs that impact the long-term planning of assets. The discussion will also touch upon the asset strategy dimensions required to achieve optimal asset performance with respect to TCO. The Total Cost of Ownership is a combined effect of all these dimensions over a period of time under consideration. The challenge is to achieve optimal assert performance based on the total cost of ownership derived using these dimensions. An important aspect of this discussion is the leverage of advancements in the technology which is a combination of computing infrastructure, instrumentation or enterprise sensory networks, and state-of-the-art methods in intelligent and knowledge based approaches. Our overall objective is to emphasize the need for use of a combination of these advancements to add value from EAM system solution's perspective.

The paper is organized as follows. Section 2 discusses the typical asset strategy levels in an enterprise and the concept of total cost of ownership. A discussion on dimensions of asset performance and prime drivers of strategic asset related decisions is covered in section 3. The need for model based approach to asset management maturity and movement of information/decisions across levels of the maturity is discussed in section 4. Role of technology in intelligent asset management solutions is emphasized in section 5. Ultimately, section 6 summarizes the key message of the discussion.

2 ASSET STRATEGY LEVELS

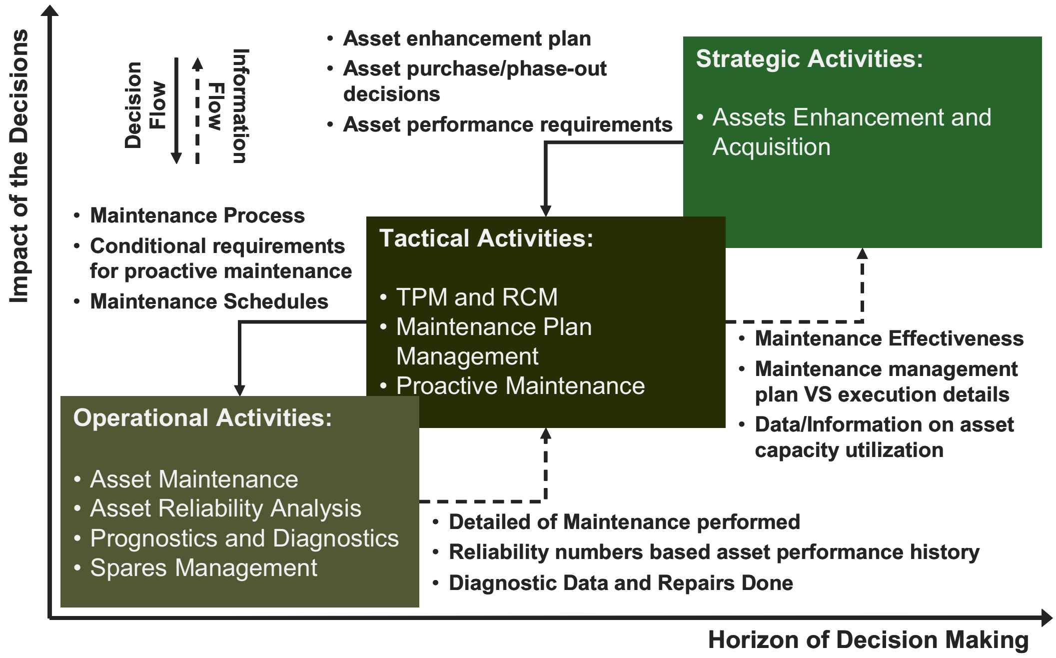


Figure 1: Asset Management Strategy Levels

Enterprise asset strategy is a combination of decisions that are at multiple levels and varying time horizons of decision making. Figure 1 shows broad categorization of the asset management activities and decisions into three levels of an enterprise.

Operational: These are a set of functions necessary to keep the asset(s) up and running to the enterprise need. Operational level aims at providing the bare minimum intelligence required to bring together all the ingredients of the maintenance function along with the diagnostic ability required in case of failures or breakdowns.

Tactical: Functions in this layer are proactive than reactive and aim at making the asset maintenance in line with overall planning. Monitoring of assets and planning of their maintenance triggered by conditions is managed by this layer. The overall process of operational asset management is driven from the tactical layer.

Strategic: The strategic layer does long-term asset planning from TCO perspective. This layer is the key in choosing amongst a number of asset acquisition options available based on the most optimal total cost of ownership. Decisions at this level are long-term and high impact as they involve capture of total ownership cost over entire asset lifecycle. The other two layers primarily work towards complying with the performance expectations set by this layer while making strategic asset related decisions for a long time horizon.

2.1 Operational

Operational layer of asset management concentrates on fulfilment of tasks related to maintenance activities of different types, capture of data/information, and analysis towards understanding the asset behaviour better. Use of the knowledge acquired at this level is brought in use by levels above it.

Asset Maintenance: Involves preventive (frequency or run-hour based), corrective (event based), and management of maintenance work force.

Reliability Analysis: Reliability of a system is defined as the trustworthiness to do what the system is expected or designed to do. At the system level the reliability can be characterized by a number of metrics such as POFOD (probability of failure on demand), ROCOF (rate of failure occurrence), MTTF (mean time to failure), and AVAIL (availability or uptime). Reliability of complex systems consisting of interconnected sub-systems needs a more comprehensive consideration owing to the individual nature of the sub-systems, behaviour of these sub-systems when interconnected with each other, effects due to propagation of faults/failures through interconnections, and effects of sub-system level interaction on higher level systems. While the systems (or complex equipments) are hierarchical by their design, this alone may not be sufficient to capture the effects of reliability of individual sub-systems on the higher-level complex system. Functional dependencies in addition to hierarchical dependencies need a careful consideration for creating sound models for equipment reliability.

Prognostics and Diagnostics: Up-keeping of assets and their health is an important aspect of operational asset performance. Being able to detect and predict forthcoming failures in the assets is a much required capability that asset management solutions need to possess to make the long term asset planning a well-informed exercise. Three primary levels of diagnostics are detection of abnormal conditions, prediction of abnormal conditions, and fault diagnosis to gain better understanding of the dependencies of these conditions. Prognostics are a proactive form of diagnostics where it is equally important to trace back to the root cause while predicting.

Spares Management: The demand for spares is triggered by periodic spare replacement as well as corrective spare replacement triggered by breakdowns. While the planning of the spares requirement is done by the tactical layer above the operational one, the key measure at this layer is making the right spare available whenever required by a maintenance task.

2.2 Tactical

The tactical layer uses the knowledge and understanding built at the operational layer to manage the operations effectively. The prime objective of this layer is to ensure that the asset management plans are in place and the operational layer indeed adheres to them.

Maintenance Plan Management: Involves a number of functions related to management of different types of maintenance activities. Creation of maintenance schedules and ensuring the availability of required resources to fulfil them is the key function here.

Proactive Maintenance (Condition or Prediction based): Monitoring of the assets and performing maintenance based on the current condition is the function also covered at the tactical layer. The idea is to avoid any major breakdowns or failures in the future and take preventive actions well in advance.

Reliability Centered Maintenance (RCM): Reliability Centered Maintenance (RCM) is a best practice approach for identifying and implementing the most effective maintenance program for any group of assets. Asset functionality is the primary focus and RCM provides a way to select the maintenance strategy that best balances maintenance costs against losses to the organization when functionality is impaired.

Total Productive Maintenance (TPM): Predictability of asset uptime and availability is an important factor in ensuring productivity of the assets. TPM works towards prevention of deterioration and not towards repairing machines/assets reactively. Key measure of TPM is Overall Equipment Availability (OEE).

Collaborative Management: The tactical layer also tries to connect with the related functions in the enterprise, such as, production, materials planning, and operations. Effort is to bring in the collaboration so as to achieve best possible utilization of the asset.

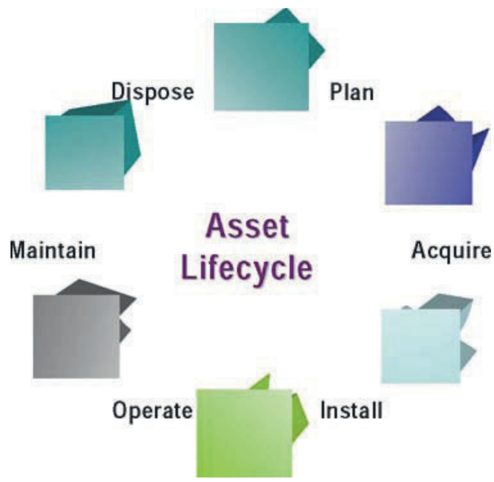


Figure 2: Typical Asset Lifecycle

2.3 Strategic

Asset management at strategic levels deals with long term planning decisions such as asset acquisition, asset phase out, and setting asset performance requirements for planning and operational layers of asset management.

Figure 2 depicts the typical asset lifecycle at strategic level of asset management. The asset lifecycle begins with asset planning and acquisition. After successful installation of the assets in operations the lifecycle continues with the operations and maintenance of these assets till it reaches the phase out stage. Primary purpose of asset lifecycle management is to maintain or increase the asset values through a series of these lifecycles. Significant decisions at the strategic level are that of acquisition of new assets or major enhancements to existing assets. The very decision between these two at any given point is also a trade-off at the strategic level of asset management. Asset operated in an optimal manner provides value and is an important aspect of asset management but that does not comply with the concept of asset optimization based on TCO. Optimization of TCO essentially involves continuous monitoring of profits and loss of each of these asset lifecycle phases and focus on asset optimization across the lifecycles [2].

2.4 Total Cost of Ownership (TCO)

Total cost of ownership (TCO) is a financial estimate of direct and indirect costs related to purchase and operation of assets over their lifecycle. TCO should ideally incorporate not only the cost of purchase but all aspects in the further use and maintenance of the equipment, device, or system under consideration. It includes the costs of training support personnel and the users of the system, costs associated with failure or outage (planned and unplanned), diminished performance incidents (waiting time), costs of security breaches (in loss of reputation and recovery costs), costs of disaster preparedness and recovery, floor space, electricity, development expenses, testing infrastructure and expenses, quality assurance, incremental growth, decommissioning, and more case specific aspects.

Many organizations fail to understand and calculate TCO and instead rely on TCA (Total Cost of Acquisition) in making new asset acquisition decisions. Also, in some cases, purchasers may not bear all of the TCO-related costs directly even though their parent organizations always do (i.e. "cost shifting"). TCO can and often does vary dramatically against TCA, although TCO is far more relevant in determining the viability of any capital asset investment. TCO also directly relates to total costs across all projects and processes and, thus, organization’s profitability.

3 ASSET PERFORMANCE ENHANCEMENT

Measurement is a key to performance enhancement. An ability to understand the current maturity and performance level of the asset management process is an essential element in being able to drive continuous performance optimization. A model based approach to drive EAM helps in defining the objectives as well as the drivers leading to those objectives.

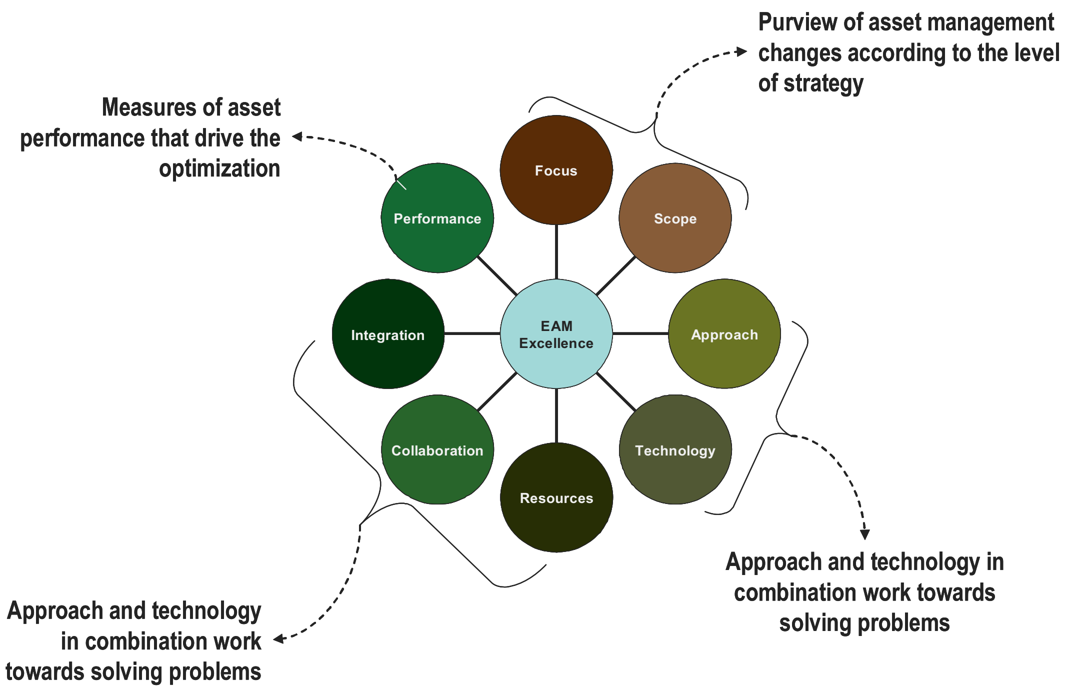


Figure 3: EAM Excellence Model Dimensions

Figure 3 shows important dimensions of the EAM excellence model (based on ARC Multi-Attribute Model of Asset Management) [3] that can be described as:

Focus: An asset can be defined as a single equipment of a large collection of equipments. Coverage of the focus dimension changes based on the level of asset management one is looking at.

Scope: Is defined based on the level whether it is operational, planning, or strategic.

Approach and Technology: Approach and technology is selected based on the focus and scope of asset management activity. This can range from well-defined processes for maintenance activity all the way to intelligent methods or formal optimization based techniques at strategic levels.

Integration, Collaboration, and Resources: These are tightly related to each other as the integration enhances the collaboration between different resources that are part of the asset management world. Information Technology based solutions with use of appropriate approach and technology can achieve this effectively.

Performance: Measurement of asset management effectiveness through well-defined Key Performance Indicators (KPI) at different levels of asset management functions.

Over 65 percent of manufacturing investments are in production facilities that directly affect revenue generation [3].

3.1 Objective Drilldown

Asset performance is defined by a number of top level objectives that need to be drilled down to requirements and activities that fulfil them. This drilldown is essential to achieve the objectives as well as facilitate the decision involving capital investments.

Figure 4 is an example of a typical drilldown from a top level objective such as "Ensuring Asset productivity". Immediate dependencies in ensuring asset productivity are asset utilization, reduction of operating costs, and prolonging of productive asset life. They further drill down into their dependencies finally leading to the set of activities and processes that need to be in the place to fulfil the objective. As can be see clearly from Figure 4 there are a number of conflicting dependencies in the hierarchy. Replacement of non-productive parts conflicts with the aim of keeping the operational costs low or the need for regular overhauling conflicts with ensuring close to 100% available time. The challenge is to weigh these actions on the basis actual cost and select the most optimal balance that would achieve the top level objective.

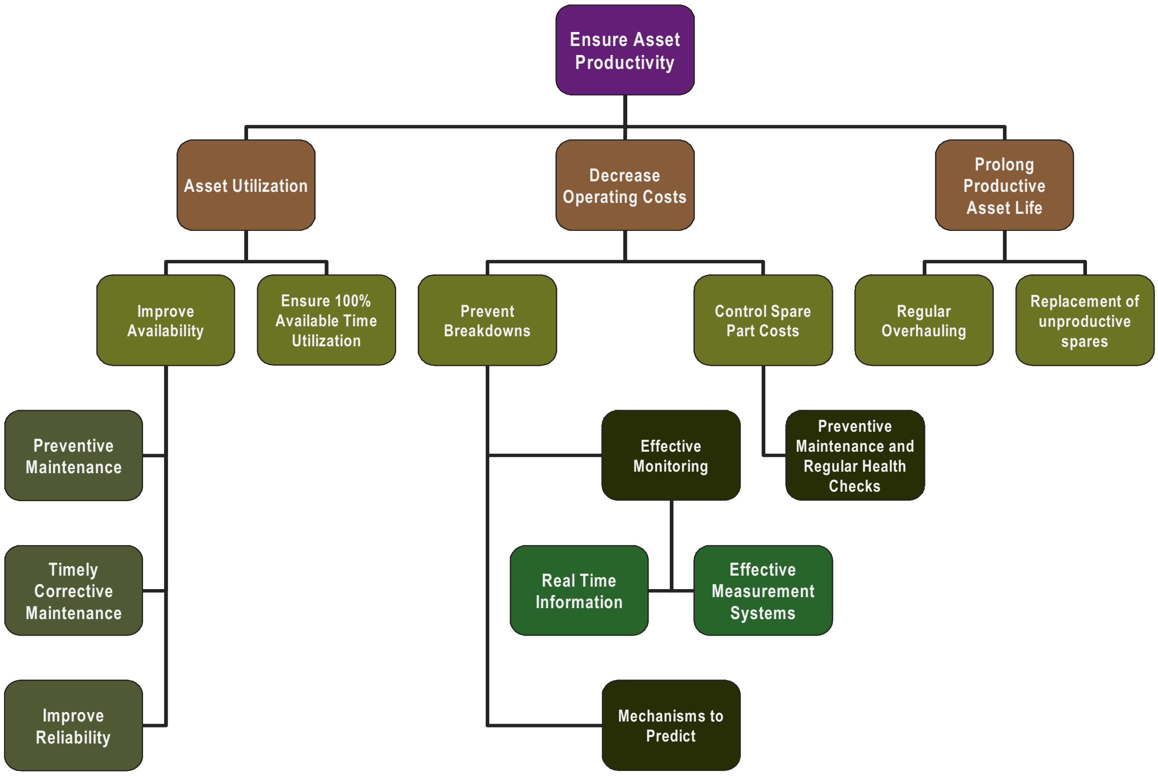


Figure 4: Typical Objective Drilldown

3.2 Capital Determinants in Strategic Decisions

Capital determinants of asset strategy are part of maintain, dispose, and plan phases of typical asset lifecycle shown in **Error! Reference source not found.**. The key understanding required is the point at which the organization should stop maintaining the existing assets and move towards acquiring enhanced newer ones by disposing the existing ones. This decision involves a number of drivers such as capital investment required to acquire new assets, operating costs of existing assets, operating costs of the assets being newly acquired, impact on productivity, reliability and residual life of the existing assets, and a number of qualitative factors, all of them leading to the strategic objective of the expectation of return on investment on the new assets. The key drivers listed above can be described in detail as:

Capital investment on new assets: Investment of capital required to acquire new assets is a major consideration in asset acquisition decisions. The investment amount in combination with the time required to return on the investment forms an important element of the go or no-go decision. Investment amount comprises of mainly the cost of assets, cost of space required to house the assets, cost of training, and skill enhancement required.

Expected operating (O&M) cost of new assets: It is not only the capital investment required but also the expected operating and maintenance cost of the new assets that adds to the consideration for new asset acquisition. Build-up of the operating costs over the initial investments in combination with the qualitative factors can have significant impact on the decisions. The operating cost changes over time and needs to be considered over entire lifecycle of the new asset.

Operating (O&M) cost of existing assets: Operating and maintenance cost of existing asset over its remaining life is also a consideration in acquisition decisions. As the assets get old the cost of maintenance and operations starts ramping up. While, this sounds trivial but its combination with the capital investment for new assets, O&M cost of new assets, and impact on productivity coupled with a number of qualitative factors is not so straightforward.

Impact on productivity: Comparison in terms of impact on productivity with induction of newly acquired assets. It is trivial that the productivity has to go up in case of new asset acquisition but the order of improvement is what matters. Productivity incorporates a number of factors such as time for maintenance, typical waiting time, percentage availability, and inherent productivity of the asset when it is running.

Reliability and residual life of existing assets: O&M costs and productivity do not remain attractive towards end of asset life. The residual life and the expected behaviour of the asset reliability over residual lifetime is important and links very closely with the overall effectiveness of the asset in terms of its availability and productivity.

Qualitative factors: A number of qualitative factors also form part of this decision making process. Aspects such as changes required in the skills of people to use the new assets, change in number of people required, current productivity agreements with worker unions, and environmental factors. Although mostly qualitative, these factors have a significant impact on the decision to acquire new assets. A number of qualitative factors may get added based on specific situations such as geography, maintenance and service support possible, maturity of new technology, etc.

A combination these costs and factors give an estimate of what can be called as the expected return on investment towards new assets. A comprehensive model for comparison of new assets with existing ones helps in supporting such decisions with analytical backup. The optimization problem becomes straightforward in case of extreme situations where the existing assets are highly unreliable or at the fag end of their productive life. It is not only a decision on whether to go for new assets or not but also decide on the most optimal time to go for them. Given the projections on various cost factors an organization needs to simulate a number of decision options in terms of exact time of transition to new assets and find out the expected time of return on investment.

Another important consideration at such juncture is the choice between buying the new assets and adopting a service oriented model for use of the assets. The service model is prominent in civil aviation industry where the cost of assets is prohibitively high both on acquisition and O&M front.

The approach to this decision needs to be model based given the need for continuation with a similar TCO based approach post acquisition. This aspect of asset lifecycle is discussed in the following section.

4 REAL-TIME INFORMATION BASED ASSET PERFORMANCE ENHANCEMENT

While the decisions about new asset acquisition are made based on TCO it is equally important to continue the use of this model throughout its lifecycle till its disposal and replacement by newer asset. The idea is to subject the TCO based model to real time asset performance information during its lifecycle and use the feedback to make changes/improvements to asset management practices within the organization. The TCO based model would essentially generate the actual TCO on real-time that can be compared with the projected TCO at the time of new asset acquisition to trigger these changes.

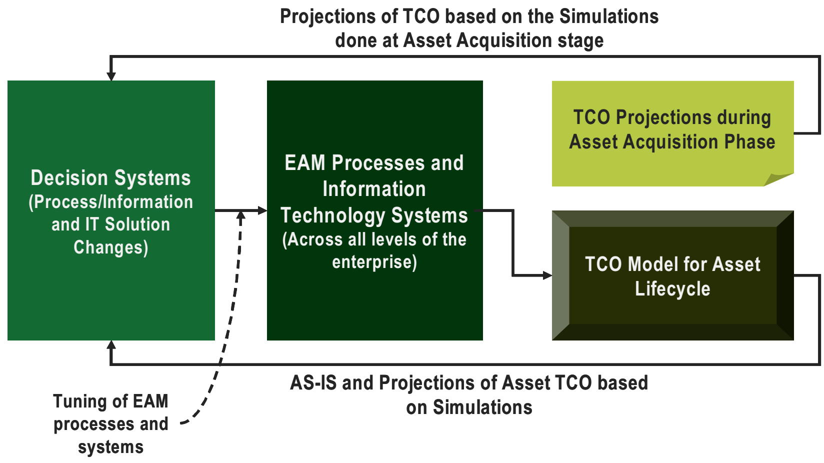


Figure 5: Feedback Based Asset Performance Enhancement

The changes to asset management processes and systems have to be drilled down the layers of the enterprise all the way to the operational layer. The decision systems to bring about these changes are also spread across the layers of the enterprise. The EAM maturity model discussed in the following section talks about the levels that an enterprise can strive to graduate across in the process of driving asset management using the TCO model.

4.1 EAM Maturity Model

Asset management is essentially a layered function that spans across definite acts of corrective maintenance all the way to strategic decisions leading to acquisition of new assets. The maturity of EAM in an organization is a chain of levels leading to topmost layer of achieving optimization on TCO basis.

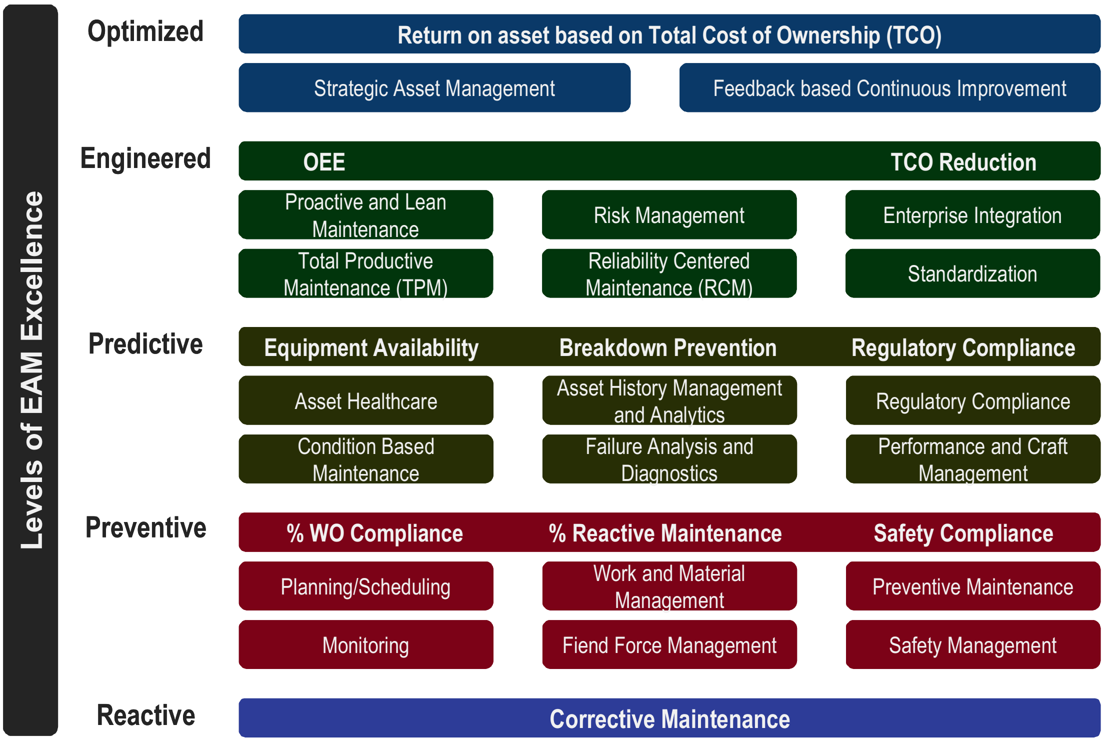


Figure 6: EAM Excellence Maturity Model

The levels of asset management maturity can be thematically described as:

Reactive: This level focuses on getting back to normal from a breakdown or failure situation at the earliest. Maturity at this level essentially means standardization of repair processes and procedures.

Preventive: Measures required in avoiding a breakdown or a failure is the prime objective of this level. Monitoring and preventive maintenance are the prime drivers to achieve the maturity required. Key performance measures for this level are defined around compliance to preventive maintenance plan, degree of reduction achievable in reactive maintenance, and compliance to safety related measures. The outcome should be to avoid an event that would result in loss of asset availability or major expenses towards recovery of the assets from breakdown.

Predictive: This is a knowledge oriented level of asset management. Use of past experience in combination with monitored data is the key to achieve predictability. The prime objective is to build knowledge using history and use it to enable the lower level of preventive asset management to avoid breakdown situations. Another important function this level does is that of continuous learning using the feedback from the lower layers. Key performance measures for this layer are defined around availability of equipment/asset achieved, number of breakdown situations prevented, and compliance to regulatory norms. Ability to learn, create knowledge, and use it for effective asset management at lower levels is the theme of predictive level of maturity.

Engineered: Complete view of total cost of asset ownership and integration across different functions of the enterprise is the theme of this layer. Total Productive Maintenance (TPM) and Reliability centred Maintenance (RCM) to achieve the near optimal asset management function over a time horizon is the desired key achievement of this layer. Overall Equipment Effectiveness (OEE) is the underlying objective to drive the activities at this layer of maturity. It also serves as a build-up for the higher layer of ensuring optimization of asset management over multiple lifecycles.

Optimized: The highest layer of asset management maturity where the key is to achieve asset optimality over multiple lifecycles. Being able to have a long term view of asset performance is the primary requirement of this layer. As discussed before the optimization of TCO involves continuous monitoring of profits and loss of asset lifecycle phases and their use in arriving at optimal decisions. Another aspect is to continuously monitor the costs and gains as the assets progress in their lifecycle to enable a feedback controlled mechanism into asset management. Performance optimization of real-life situations cannot be achieved without feedback and this crucial ability is required at the optimized layer.

The EAM maturity model [5] layers map to the dimensions shown in Figure 3 as given in Table 1.

Table 1: Maturity Mapping to Model Dimensions

Dimensions	**Asset Management Maturity Level**				
	Reactive	Preventive	Predictive	Engineered	Optimized
Focus	Machine	Machine	System	Process	Enterprise
Scope	Fault Correction	Reliability	Performance Improvement	Functionally Improvement	Enhancements and New Acquisitions
Approach	Corrective	Preventive	Predictive	Engineering	TCO Optimization
Technology	NONE	Usage Monitoring	Health Monitoring	Online Health Monitoring and Decisions	Model based optimization and self-diagnostics
Resource	Plant	Plant	Business	Enterprise	Asset Lifecycle
Collaboration	Maintenance Work Force	Planning and Scheduling	Diagnostic Knowledge	Decision Support through defined goals	Decision Support through well defined responsibility
Integration	Spares Management	Procurement	Production Management	Enterprise	Supply Chains
Performance	Service Efficiency	Asset Availability	Asset Performance	OEE	Return on Assets (ROA)

4.2 Information Feedback

The maturity model is not complete without a well defined information and decision flow across the maturity layers. The model maturity layers essentially translate into processes or IT based systems to automate the processes at every level. The flow of information upwards into the maturity layers enables decision making to strategically drive the layers below.

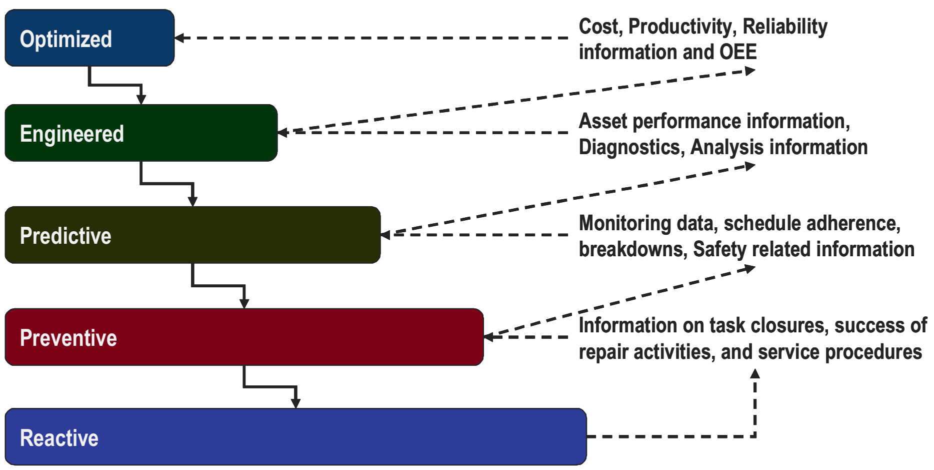


Figure 7: Information Feedback up the Maturity Layers

Technology plays an important role in establishing the connectivity between the systems and processes at each level of maturity. While the maturity model is conceptual the EAM systems has to be a combination of applications working at various levels to achieve the asset management maturity. The role of technology is to enable measurements, integration, and facilitate decision making through use of advanced algorithms and methods in optimization. Next sections discusses how recent the advances in technology have made this possible is covered in the next section.

5 ROLE OF TECHNOLOGY IN INTELLIGENT ASSET MANAGEMENT

Evolution of advanced technology in recent years had contributed positively to deployment of intelligent decision-making solutions in industry. Asset management is not an exception and continues to use the technological enhancements to its advantage. Three significant dimensions of technological advancements are:

Computing: Speed of computing has considerably increased over past decade and that has led to more widespread use of high speed computing in applications as never before. The increase in computing speed is also coupled with reduction in cost and size enabling use of powerful computers in applications closer to manufacturing layers of an enterprise. The standard PC environment is now capable of running heavy number and data crunching applications in short times leading to use of intelligent decision making algorithms more commonly.

Instrumentation and Sensing: Availability of information is the key to effective decision making process. Development of technology to realize sensory networks that provide complete information about the assets on real-time is possible because of availability of advanced instrumentation that can span across an enterprise. With monitoring going real-time it is possible to combine the real-time information with high computing power as enablers for use of advanced methods in decision support.

Advanced Algorithms: Learning from data and use of intelligent algorithms to arrive at real-time decisions can be achieved by using advanced algorithms given the technological advanced mentioned above. It is possible to use formal methods in multi-objective optimization to arrive at decisions that are based on TCO. Such problems can be large scale and complex by nature but availability of high-speed computing infrastructure and real-time sensory information makes it possible to tackle them in finite time.

5.1 Technology Investments

Setting up information movement and decision support mechanisms across the layers of asset management systems of the enterprise necessitates investment towards a variety of enabling technologies. The enablers can be classified into the following:

Sensory Mechanisms: Online measurements are the key to process improvement. Measurement of operational parameters of the assets are required for condition based maintenance, assessing asset reliability, asset health management, diagnostics, knowledge capture for intelligent maintenance, and number of other decision making mechanisms in the higher levels. Sensory mechanisms are either hardware or software based. Software based methods are necessary in the situations where the operating parameter has to be estimated rather than measured. The need is to create a network of sensors that capture complete snapshot of operating condition of the assets at any given time.

Instrumentation and Historian Systems: While the measurements are used online for condition monitoring, a greater use is in capture of knowledge and analysis to facilitate decision making at higher levels. Instrumentation systems to connect the sensory mechanisms to the higher layers and historian systems to maintain history of the asset operations are required at this level. Availability of operating data of the assets over their lifecycle is the key to asset performance analysis.

Knowledge Capture Mechanisms: The operational data is a goldmine for learning more above asset behaviour and organization need to invest towards acquiring knowledge extraction algorithms that do this job. The knowledge capture is along the application such as condition based maintenance, reliability analysis, root cause analysis, prediction mechanisms, and residual life analysis. A number of methodologies such as data mining, text mining, learning algorithms (e.g. ANNs), statistical analysis, correlation analysis, and case based learning are available for knowledge capture. The key is to form an effective combination of these methods to facilitate capture of knowledge in the form that is usable by higher layers of asset management function.

Maintenance Management Systems: Commonly known as Computerized Maintenance Management Systems (CMMS) that take care of the operational maintenance requirement for the organization. Majority of the operational functions at the reactive, preventive, and predictive layer are handled by CMMS.

Optimization Tools: Formal methods in optimization are essential for trade-off analysis along a number of cost dimensions in decision making. As discussed in section 3.2, these trade-offs are almost never obvious and need a comprehensive mathematical model along with optimization algorithms to solve it using real cost data.

Decision Support Systems: Results of optimization can never deliver the final conclusion at the strategic levels. Optimization serves as a tool to compare number of options available and reduce them to finitely many to facilitate quick decision making. A number of qualitative factors then play an important role in strategic decision making. This has been discussed before in section 3.2.

Apart from these investments, a significant investment may be required on the infrastructure front. Computing and networking hardware are primary components of infrastructural investment. Their discussion is out of scope in this context.

6 SUMMARY

This discussion is aimed at three important messages related to effective asset management. First is the significance of Total Cost of Ownership (TCO) and its use in not only in acquiring the assets but also in running the assets in an optimal manner. Second is the importance of model based approach to achieving the maturity of asset management function in an enterprise and third is the need of advanced technology in achieving the information movement and decision support mechanisms required in place to achieve the maturity of asset management along five levels of the maturity model discussed. The success of strategic asset management decisions is a combination of asset management process maturity, information feedback, and use of the feedback for ongoing decisions across multiple lifecycles of the assets.

7 REFERENCES

[1] Sid Snitkin, "A Multi-Dimensional Model for Assessing Your Asset Management Program," ARC INSIGHTS, INSIGHT# 2005-04EM, January 26, 2005

[2] Houghton LeRoy, " The Future of EAM," ARC White Paper, ARC Advisory Group, November 2001

[3] Sid Snitkin, "Collaborative EAM Achieves the Full Vision of Enterprise Asset Management, ARC White Paper, ARC Advisory Group, November 2004

[4] Houghton LeRoy and Sid Snitkin, "Dynamic Reliability Centered Maintenance," ARC INSIGHTS, INSIGHT# 2004-26EC, June 17, 2004

[5] Gautam Sardar, Ramachandran N., and Ravi Gopinath, "Model Based Approach to EAM," TCS Internal Report, January 2006.

AN AUDITING PROCESS TO IMPROVE ASSET MANAGEMENT PERFORMANCE

Visser, JK [a] and Mollentze, JJ [a]

[a] *Department of Engineering and Technology Management, University of Pretoria, South Africa.*

Abstract: An organisation's maturity in asset management practices and processes can be determined through the application of an asset management audit. The main objective of this study was to develop and validate an asset management auditing methodology. This methodology includes the development of a program for asset management improvement and the application of business process re-engineering (BPR) as a tool to realise the improvement opportunities identified through the audit. The application of an asset management audit, according to the developed methodology, can create a "roadmap" to Asset Management improvement. By optimising the asset management practices and processes, a positive contribution can be made to the profitability of any business enterprise.

Key Words: Asset management, performance, audit process,

1 BACKGROUND ON ASSET MANAGEMENT

Asset management (AM) as a concept evolved from maintenance management to provide a more holistic approach to managing the total life of a physical asset. Asset management in organisations has become the focal point to improve and optimise the contribution of a physical asset to the overall success and profitability of the organisation. In order to achieve optimum return on investment from a physical asset and maximise its contribution to the revenue or success of an organisation, the organisation needs to grow its maturity in asset management. This means that, taking into consideration the stages in an asset's life cycle, maturity levels have to be defined. An assessment is needed to map the level of maturity of the organisation's current practices and processes. Kaizer [1] states that management audits provide a framework for organisations to systematically review, analyse and recommend improvements in performance. Its purpose is also to ensure that management is carrying out its mission, meeting its goals and objectives, following proper procedures and managing resources effectively and efficiently.

2 AUDITING ASSET MANAGEMENT

Traditional maintenance management audits are conducted through interviewing key role players in the organisation with a standard list of questions, performing walk-abouts and studying some requested documents and key measures. The results of such audits are then given in the form of a score with areas of concern highlighted. Coetzee [2] says that measurement, as part of the maintenance cycle, consists of regular (typically annual) maintenance audits and a maintenance performance measurement process (typically performed monthly). The objective of the audit is to determine how well the processes succeed in achieving the results envisaged by management in setting up the maintenance policies and procedures.

2.1 Research Objectives

Many definitions for asset management have been formulated. A consolidation of the different views and opinions was needed to provide an acceptable definition for asset management. Furthermore, the framework for asset management has yet to be fully developed. Various organisations and agencies have developed frameworks which are not fully comprehensive and do not take into consideration all issues in the concept of asset management. Within the elements of such a model or framework, the basic business processes needed have not been formally identified with their desired attributes. A methodology to conduct asset management audits needs to be formulated in order to map the current performance or maturity of an organisation. Included in this methodology should be the improvement process at completion of such an audit.

The main objective of the study was to develop and validate an asset management auditing process or methodology. This has not been addressed in the literature yet. Other objectives, supporting the main objective, were the development of:

- A clear definition of asset management.
- A comprehensive, all-inclusive framework for asset management.
- Business processes associated with each element of the framework.
- An asset management improvement strategy and program.

3 LITERATURE REVIEW

3.1 Background

A literature survey was done on asset management definitions and frameworks as well as on asset and maintenance auditing methodologies. A few of the important literature sources are mentioned briefly in this paper. A detailed discussion of the literature is given by Mollentze [3].

3.2 Definition of Asset Management

Campbell [4] states that maintenance is one of the steps (or phases) in a nine-step asset management process. This process is initiated with an 'asset strategy' phase during which the question is answered on why the asset is required by the business enterprise. The process is concluded when the asset is terminated or 'phased out'.

Clarke [5] states that physical asset management involves the management, co-ordination and execution of all the business activities and processes required through the physical asset's life-cycle, from identification of the business need for a physical asset, through the acquisition, operation, maintenance and eventual disposal.

According to Woodhouse [6], asset management is the set of processes, tools, performance measures and shared understanding that glues the individual improvements or activities, related to assets, together. Woodhouse [6] defines physical asset management as *'The set of disciplines, methods, procedures and tools to optimize the whole life business impact of costs, performance and risk exposures (associated with the availability, efficiency, quality, longevity and regulatory/safety/environmental compliance) of the company's physical assets'.*

Asset Capability Management (ACM) consultants [7] state that maintenance must be defined and managed as a process. Processes need to be established in order to effectively manage maintenance activity and overall asset life cycle management.

Peterson [8] says that strategic asset management involves managing the capital investment towards a long-term program of increasing the return on assets. Peterson [8] defines asset management as *'A global management process through which we consistently make and execute the highest value decisions about the use and care of our assets.'*

3.3 Asset Management Frameworks or Models

Campbell [4] explains that before an organisation embarks on an improvement plan it should assess the strengths and weaknesses of the present system. The diagnostic must be a clear roadmap of the next steps to achieve their vision and should be comprehensive and cover strategic, procedural, technical, administrative and cultural issues. Campbell [4] proposed ten major areas that should be included in an asset management framework.

The New South Wales Government [9] formulated a basic framework for the development of an asset strategy. This framework, illustrated in figure 1, considers each characteristic as a "gate" through which assets are analysed.

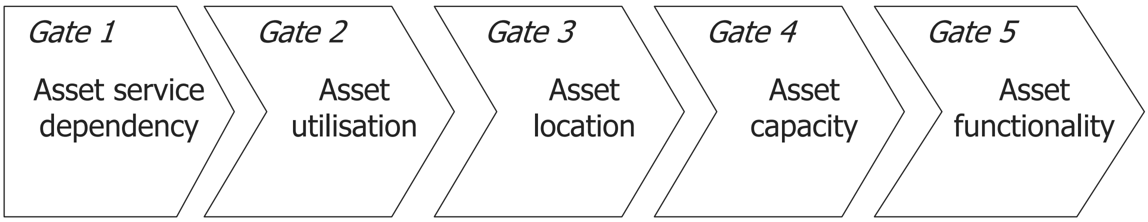


Figure 1. NSW Government framework for development of an asset strategy [8]

Woodhouse [6] views asset management as a set of integrated processes that comprise mainly resource control, work control, cost control and change control.

3.4 Auditing Methodologies

Kaizer [1] defined a seven-phase auditing methodology for maintenance management. These phases are:

- Establish priorities and schedule the audit
- Define and organize the audit
- Form the audit team
- Perform the audit
- Prepare the report

- Define management action
- Final follow-up of results

Hoberg and Rudnick [10] also defined a number of steps for an assessment/audit. Aspects like defining the scope of the audit, planning the activities, forming the team, training of team members and documenting the outcomes of the audit are mentioned.

4 THEORETICAL FRAMEWORK FOR ASSET MANAGEMENT

4.1 Definition of Asset Management

Asset Management is defined in this paper as 'The management of physical assets through its acquisition, operation, maintenance and disposal life cycle phases in order to optimise life cycle costs and maximise contribution to the profitability or success of an organisation'.

4.2 Asset Management Framework

According to the definition of Asset Management given above, the new high-level framework will include the four phases of the asset's life cycle. This gives a holistic scope for the concept of asset management and clearly shows that maintenance management is only one of the elements of asset management. A high-level framework for asset management, based on the four life-cycle phases, is illustrated in figure 2.

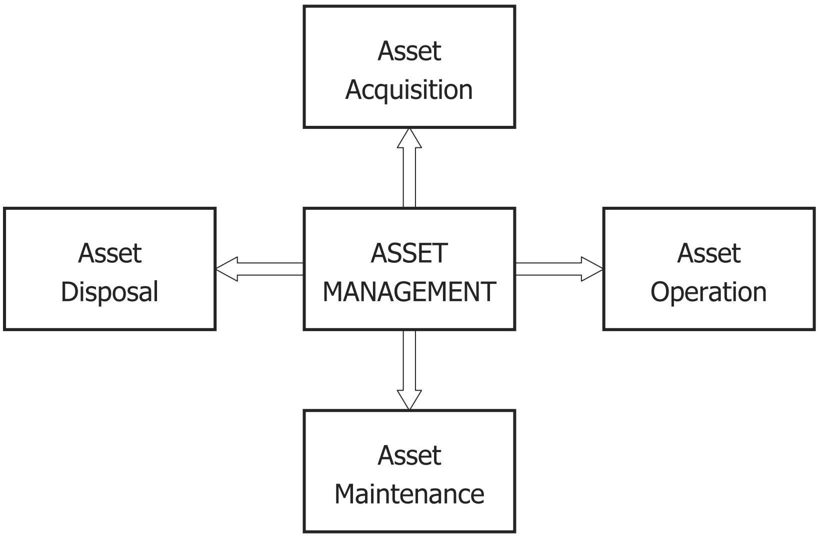


Figure 2. High level framework for Asset Management

The elements of each of the four life cycle phases of an asset are mentioned below. The characteristics of each of these elements are discussed in detail by Mollentze [3].

4.2.1 Asset Acquisition

The *'Asset Acquisition'* phase of a physical asset's life can be divided into four elements namely:

- Asset requirements analysis
- Asset design/procurement
- Asset life cycle analysis
- Asset installation and commissioning

4.2.2 Asset Operation and Maintenance

The *'Operation'* and *'Maintenance'* phases indicated in figure 2 are grouped together since they occur in the same time period. The '*Asset Operation and Maintenance'* phase comprises fifteen sub-phases or elements.

- Strategy development
- Maintenance tactics
- Work management
- Human resources
- Performance measurement
- Continuous improvement
- Financial management
- Materials management
- Systems and information
- Change management
- Outsourcing
- Safety, health, environment and risk (SHER)
- Control and automation
- Operating procedures
- Technology management

4.2.3 Asset Disposal

The final phase in the life cycle of an asset, i.e. '*Asset Disposal*', comprises the following 4 elements that represent the business processes that must be managed.

- Asset decommissioning
- Asset divesting
- Asset retrospect
- Safety, Health, Environment and Risk (SHER)

4.3 Proposed Asset Management Auditing Methodology

A number of auditing methodologies were mentioned in section 2.4. A more comprehensive methodology that includes the concept of Business Process Engineering was formulated and is shown in figure 3. As part of the audit methodology, business processes are identified, mapped and assessed according to the developed framework for Asset Management. The proposed methodology is applicable to a large plant or site and should be scaled down for smaller or sites. A brief description of each phase or element of the proposed methodology is given below. A more detailed discussion is provided by Mollentze [3].

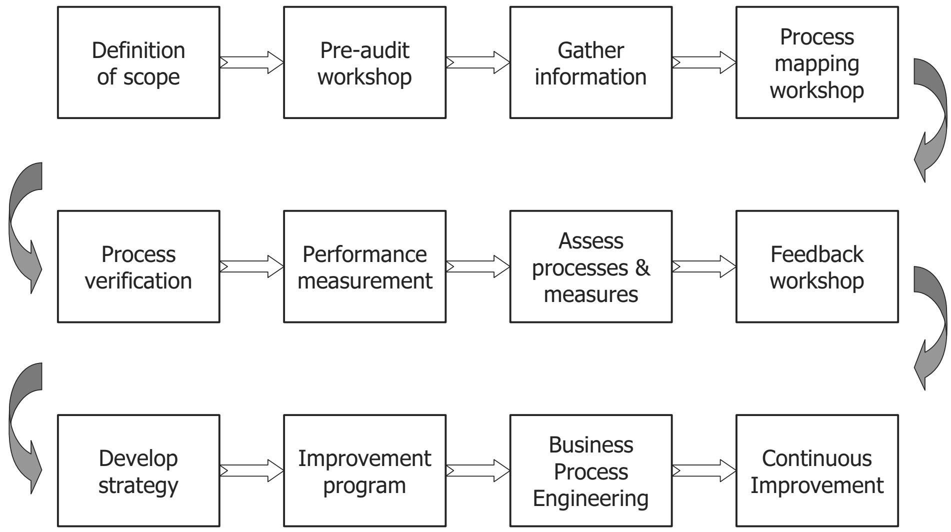


Figure 3. Proposed asset management auditing methodology

Sponsor/Management Definition of Scope

Management or the identified sponsor defines the scope of the audit together with the auditor. The framework, as discussed earlier, consists of elements and a decision has to be made whether certain of these elements will be excluded or not.

Pre-Audit Workshop

Before the audit is conducted a pre-audit workshop is conducted with the key role-players from the section. It is important to include representatives from all affected departments in the organisation, particularly operations and maintenance.

Information Gathering

According to the selected scope, information on the relevant elements is gathered. This information includes the current processes in each element in the form of policies or procedures as well as the supporting systems and tools.

Process Mapping Workshop

This step is different from the ordinary run-of-the-mill audits. After the information has been gathered and the auditor identifies that there are certain processes that he could not be supplied with (because it has never been documented), he should facilitate a short workshop to map the outstanding processes. It may also be found that the required process does not exist.

Process Verification

This can be compared to the physical audit of "traditional" audits. The auditor will visit the section and question certain key personnel on the execution of the processes in the workplace. One may find that a business process has been documented in the form of a procedure, governing the execution thereof, but that in the day-to-day situation in the workplace it is executed differently from what is prescribed.

Performance Measurement

Organisations and even sections within organisations have different ways of calculating the same measurement. The different elements of the audit framework should have a least one or two measures assigned to it to enable comparison to benchmarks.

Assessment of Processes and Measures

After the auditor has gathered all the information required he will be able to assess and appraise the organisation's level of maturity in each of the elements. A guideline for assessing the business process maturity is shown in table 1.

Table 1

Five levels of maturity for business processes in asset management

Level 1	Level 2	Level 3	Level 4	Level 5
The business process (BP) does not exist	Business process exists, not documented, no training in BP's, not been improved in previous few years	Business process exists, is documented and communicated in form of policy or procedure, readily available and understandable. Some revisions done in previous year.	Business Process training is performed, roles and responsibilities of individuals according to it. Systems support 80% of BP's.	Business Process has been modelled with ongoing BP Improvement initiative. Integrated across functional and departmental boundaries. All elements in AM covered by BP's.

Examples of some KPI's and benchmarks that can be used are shown in table 2.

Table 2

Five levels of maturity for performance measures in Asset Management

Element	Metric	Level 1	Level 2	Level 3	Level 4	Level 5
Maintenance Tactics	Planned vs. Unplanned maintenance hours ratio	40:60	50:50	60:40	70:30	80:20
Materials Management	Stores value vs. Equipment replacement value	4,0%	3,5%	3,0%	2,5%	2,0%
Financial Management	Maintenance cost vs. Equipment replacement value	8,0%	6,5%	5,0%	3,5%	2,0%
Work Management	Labour utilisation (Wrench time)	45%	50%	55%	60%	65%

According to the findings of the auditor, the maturity of performance and practices of the organisation can be mapped on these matrices. This allows for scoring and determining the average level of maturity in each element and asset management as a whole.

Feedback Workshop

The feedback workshop should include all the key personnel that were involved in the audit as well as a representative from different levels and departments. A full report can also be distributed during this session.

Strategy Development

When developing the asset management strategy, representation from all key stakeholders is imperative. This is important to ensure that the necessary support and buy-in is gained from the people that are going to make it work. Opportunities for improvement have been identified by the audit and according to this, the strategy will be developed. Aspects to be addressed in this step are the development of a vision, mission, key objectives and key performance indicators (KPI's).

Improvement Program

The gaps have been identified, targets have been set and the *Program for Asset Management Improvement* (PAMI) has to be developed. This program will contain the plans to close the gaps and achieve the targets. PAMI is a living program that details all the necessary actions, resources and timelines that can take the organisation to the desired future state. PAMI needs to be communicated and displayed to all stakeholders so that current and future actions can be visible to all concerned.

Business Process Engineering (BPE/BPR)

Tritter [11] postulates that the entire organisation can be seen as a set of processes. Various tools and methodologies are available in industry to perform Business Process Re-/Engineering. Examples include the *Business Process Re-design* of Malhotra [12], *Control Self-assessments* by Tritter [11] and *Process Reengineering Life Cycle* by Grover [13], to name a few. The organisation will have to review the methodologies available to perform Business Process Re-engineering and adopt what is most suited for the organisation.

Continuous Improvement

The Asset Management audit is not a once off exercise. It is advisable to perform follow up audits to assess the implementation of the improvement actions as well as the effectiveness thereof. As you apply the audit process and compare trends with past audits, you gain increasing knowledge about the dynamics of your organisation.

5 VALIDATION METHODOLOGY

The validation of the asset management framework and audit methodology was done by means of a practical application as well as a survey distributed to different representatives from a wide range of organisations.

5.1 Data Collection

5.1.1 Practical Application

The auditing methodology as discussed in section 3.3 of this paper was applied on the Final Recovery Plant at the Jwaneng Diamond Mine in Botswana. The first step was to determine the scope of the audit. Members of the management team at the plant decided, due to several reasons, to limit the scope to only eight elements of the asset management framework as discussed in section 3.2.

5.1.2 Survey

A questionnaire was drawn up to test the asset management framework and audit methodology in the field. The questionnaire comprised five statements and respondents were requested to indicate whether they (i) fully agreed, (ii) partially agreed or (iii) disagreed. The questionnaire was distributed within the De Beers group of companies in Southern Africa and addressed to persons directly involved in the field of asset management.

5.2 Data Analysis

5.2.1 Survey

The results of the survey are summarised in table 3 below.

Table 3

Summary of responses to the questionnaire

Question	Agree (%)	Partially agree (%)	Disagree (%)	Some comments of respondents
The definition for Asset Management as quoted in 3.1 of the article describes the concept of physical Asset Management clearly, considering the whole life cycle of assets	100	0	0	Operation and maintenance are two different processes but they occur in the same time frame in the life cycle of physical assets.
The elements of the developed framework, for Asset Management, address most of the possible elements/business processes within a typical organization as described in 3.2	57	43	0	Would prefer to see a fifth dimension, that of operational readiness, an asset taken through from being acquired to operational phase.
The framework is simplistic and easy to understand (down to Supervisor level)	57	43	0	A study of the concepts involved will have to be carried out to fully understand their meaning and the language could be simplified.
The asset management audit methodology described in paragraph 3.3 is practical and could easily be applied to your organisation	43	57	0	Asset Management should have a strong business orientation and, whatever framework is used, should support Business Value Add within an Asset Life-cycle context.
The application of the asset management audit methodology should have the potential to positively affect the profitability or success of an organisation	86	14	0	Auditing is a management tool addressing the need for assurance, much more than it is a strategy implementation vehicle. It cannot circumvent investment in people etc. but admittedly enhance the success of these investments by assuring basics are and stay in place.

The results from the survey indicate that the new auditing methodology is accepted and supported, as well as the framework and elements identified as the scope of Asset Management. Some of the comments can be used to update the audit methodology. Most of the partial agreements could not be validated as it seemed that the respondent either did not understand, or misinterpreted, the information provided before the survey was done. The biggest concern highlighted by participants was that the audit might be too holistic, but it is made clear that the first step of the audit is to determine and/or narrow down the scope. Obviously, when this step is performed the environment where the audit will take place must be taken into consideration and this will ensure that the scope of the audit is then applicable for the specific application and is not too broad.

5.2.2 Practical Application

The auditing methodology described in this paper was applied successfully at the Red Area, Jwaneng Mine and the stakeholders of the Red Area supported the new auditing methodology and were willing to use it on a regular basis. The results from the audit were regarded as accurate and supported their subjective evaluation of the asset management function.

6 CONCLUSIONS AND RECOMMENDATIONS

The main objective of this research, i.e. to formulate a comprehensive framework for asset management, and to develop and validate an auditing methodology, was achieved. Aspects of the Business Process Re-engineering methodology were included in the audit methodology.

The proposed audit methodology can be applied in virtually any business enterprise. It was tested practically at a section of the Jwaneng diamond mine in Botswana and the results were satisfactory for the mine management. Several opportunities were identified within the predetermined scope of the audit at the site that otherwise may not have been identified. The Program for Asset Management Improvement will contribute to improving the performance of the production assets.

A contribution was also made to the theory of asset management through the development of a comprehensive framework for asset management. The framework and elements of the framework contribute to the theory of asset management by defining the various issues that should be addressed through asset management during the life cycle of a physical asset.

7 REFERENCES

[1] Kaizer, HH. (1991) *Maintenance Management Audit*, Construction Consultants and Publishers

[2] Coetzee, JL. (1999) An holistic approach to the maintenance problem, *Journal of Quality in Maintenance Engineering*, **5**, (No. 3), p 276-280.

[3] Mollentze, FJ. (2005) Asset *Management Auditing – The Roadmap To Asset Management Excellence*, MSc Dissertation, Department of Engineering and Technology Management, University of Pretoria, South Africa.

[4] Campbell, JD. (1999) *Uptime: Strategies for Excellence in Maintenance Management*, Productivity Press, p10-20

[5] Clarke, P. (2002) Physical Asset Management Strategy Development and Implementation, *Proceedings of the International Conference of Maintenance Societies (ICOMS)*, May 2002, Perth.

[6] Woodhouse, J. (2004), Asset Management – An Introduction, Institute of Asset Management (IAM), http://www.iam-uk.org/iam_publications.htm, Accessed on 07 January 2004.

[7] Asset Capability Management (2004) Asset Management – A source of additional profitability, http://www.assetcapability.com.au/articles/pace_april02.pdf, Accessed on 09 January 2004.

[8] Peterson, S.B. Developing an Asset Management Strategy, Strategic Asset Management Inc, http://www.samicorp.com/Publications.html, Accessed on 19 May 2004.

[9] NSW Government Asset Management Committee, Total Asset Management Manual, http://www.gamc.nsw.gov.au/tam/, Accessed on 01 December 2003.

[10] Hoberg, WA & Rudick, MF. (1994) *The role of assessments in a switching supplier's TQM system*, *IEEE Global Telecommunications Conference*, **3**, p 1591-1595.

[11] Tritter, RP. (2000) *Control self-assessments: a guide to facilitation based consulting*, John Wiley & Sons, p 61-72.

[12] Malhotra, Y. (1998) Business Process Redesign: An Overview, *IEEE Engineering Management Review*, **26**, (No. 3).

[13] Grover, V and Kettinger, WJ. (1995) *Business process change: re-engineering concepts, method, and technologies*, Idea Group Publishing, p 208-401.

RESEARCH ON INFLUENCES ON MAINTENANCE MANAGEMENT EFFECTIVENESS

L Gouws[a] & J Trevelyan[b]

[a]*PhD Research Student, School of Mechanical Engineering, M050, The University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009*

[b]*Professor in Engineering, School of Mechanical Engineering, M050, The University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009*

Abstract: Maintenance management is a complex partnership between the people directly or indirectly involved in maintenance, the processes and strategies to direct the people's actions, the physical assets being maintained, and the Computerised Maintenance Management Systems - the main link between the "people", the "processes and strategies", and the "asset". Although the maintenance literature contains many recipes for success, anecdotal evidence suggests that maintenance management is often ineffective. Despite advances in instrumentation, computer capabilities, integrated asset management software, and the involvement of reliability engineers and maintenance managers, unexpected and costly equipment failures still occur – sometimes with adverse effects on people and the environment. Literature on systematic research to support or contradict the anecdotal evidence is, however, hard to find. This paper introduces a research project which aims to identify and analyse various influences on maintenance management. The research (using contemporary social sciences investigation methods) is conducted by experienced engineers, and is part of a larger project on the nature of engineering work. The paper also provides a literature overview aimed at providing an insight into the major aspects leading to successes or failures in the field of maintenance management.

Key Words: SAP™ PM, Plant Maintenance, Computerised Maintenance Management System

1 INTRODUCTION AND INITIAL PROBLEM STATEMENT

Maintenance is a huge industry. On a global scale, Thomas [[1]] estimated maintenance cost at 9% of an estimated world Gross Domestic Product (GDP) of 45,000 billion AUD. Not only is maintenance on global scale huge, but individual companies also experience maintenance costs as substantial. Hägerby and Johansson [[2]] reported that the total cost of maintenance for the companies reviewed by them, is as high as 40% of company turnover.

The size of the maintenance industry resulted in the abundant supply of software, instrumentation, analysis techniques and specialists to support the maintenance industry. Landmark books have been published in the last decade that contain "recipes" to optimise companies' maintenance systems and procedures. Examples of techniques widely used by companies in projects aimed at optimising their maintenance costs include Reliability Centred Maintenance (RCM) [[4]], root cause failure analysis [[5]] and preventive and predictive maintenance (PM) [[6]]. Furthermore, our understanding of functional failures of equipment and how these present themselves in an operational environment has improved dramatically; and we now have generic preventive maintenance tasks and typical maintenance intervals that can be used as a main source when developing preventive maintenance programs. There have been dramatic advances in computer capabilities in the last decade; and instrumentation that monitors the condition of equipment has also become better and cheaper. The larger companies have maintenance engineers to address the operational day-to-day maintenance issues, and reliability engineers to address long term maintenance strategies.

Despite all of this, anecdotal evidence suggests that maintenance management still has huge unsolved issues, and that improvements in the last decade have not been as beneficial as one would expect. Unexpected and costly failures are still experienced - sometimes with far reaching consequences and ramifications. An example of such a recent event linked to maintenance was a burst main water pipe next to a freeway in Perth, Western Australia. An estimated 12 megalitres of water flooded the neighbouring areas in a short period, resulting in the hasty evacuation of business premises and private homes. Vehicles were stuck on the highway and major roads for several hours, ambulances could not transfer patients to hospitals, parents could not pick up their kids from school, taxis could not get rid of their passengers and people missed their flights.

Very little systematic research could be found on whether the anecdotal evidence about the efficiency and effectiveness of maintenance is true or merely the perceptions of a small group of individuals. If the perception that maintenance still has huge issues is shared by a larger group of people in the maintenance field, no clear reasons for this have as yet been identified. The research literature found on maintenance management mainly focuses on technical issues and advanced analysis techniques. The literature is also predominantly prescriptive. Many writers have contributed suggestions and techniques for effective maintenance programs, but these are either based on their own experiences or are aimed at promoting a specific product.

2 RESEARCH PROJECT AIM

This paper introduces a research project aimed at beginning to identify common issues experienced in the field of maintenance management, and how these issues are being (or have been) addressed by the companies participating in the research project. The research, using contemporary social sciences investigation methods, is conducted by experienced engineers and is part of a larger research project at the University of Western Australia into the nature of engineering work [[3]].

Because of the uncertainty on where the real issues are, it was decided to develop a list of possible explanations for the problems experienced in maintenance management. This list is based on information extracted from various articles, past experience of members on the research team, and from discussions with various people involved in different aspects of maintenance. Section 4 of this paper provides more details on this potential list of issues (referred to as "possible hypotheses" in the remainder of the paper).

The project will be undertaken in two phases, with the main aim of the first phase to gather real data on what people in maintenance roles experience as current issues in maintenance management. This data will be analysed, and the initial list of very broad potential hypotheses adapted to only those that seem to contribute to the problems in maintenance management. The second phase will analyse in more detail only those common experiences that emerged during the fist phase.

The main source of data for both phases will be interviews with people from a wide range of maintenance roles. Where considered appropriate, data extracted from analysis of the interviews will be supplemented with "shadowing" (i.e. following) of some people, interrogation of the Computerised Maintenance Management System (CMMS) and a review of relevant company documentation. As a minimum, it is planned to interview maintenance engineers, reliability engineers, maintenance technicians and operators. If time allows and if more information is required, inspectors, purchase buyers and accountants will also be interviewed. All interviews will be analysed using well-established qualitative analysis techniques. These techniques provide guidelines on the development of the questionnaires, as well as the coding and interpretation of the data.

3 DEFINITION OF MAINTENANCE MANAGEMENT

In the context of this research project ***maintenance management*** is defined as the preserving of the ability of an asset to safely and economically produce something through structured management of the following elements throughout the life cycle of the asset:

- The asset - the physical items requiring maintenance (e.g. plant equipment, fleets and buildings).
- People – the individuals in the organisational structure involved directly or indirectly with maintenance. Some people are very visible in the maintenance work flow process (such as asset managers, maintenance engineers, maintenance supervisors, maintenance technicians and plant operators) while others are less visible, but not less important, in the maintenance work flow process (e.g. reliability engineers, inspectors, accountants, purchase buyers, and CMMS Administrators).
- Processes and Strategies – defined as the maintenance-related business processes, maintenance strategies implemented to optimise maintenance cost, maintenance engineering and analyses techniques.
- Computerised Maintenance Management Systems (CMMS) – the heart of each modern maintenance system, and the main link between the "people", "processes and strategies" and the "asset".

4 POSSIBLE HYPOTHESES TO BE ANALYSED IN FIRST PHASE OF PROJECT

The remainder of this section contains a list of possible hypotheses that might explain some of the problems experienced in maintenance management. These hypotheses address the following three of the four elements of maintenance management as defined in section 3:

- The people
- Processes and strategies
- The CMMS

The asset being maintained is not included in the scope of this research project.

4.1 Hypothesis 1 related to "People"

"Maintenance" has negative connotations and as a profession it has little social status

The following saying by an experienced consulting engineer was once overheard "People don't stand around the barbecue bragging to their friends that they had a meeting with their maintenance engineer. However, they like telling others about their accountants, lawyers and medical specialists". It seems that maintenance is not seen as a "sexy" job when compared with other professions, and even with engineering in general. Even the word "maintenance" seems to have negative connotations. When Palarchio [[7]] for example asked a group of managers from maintenance and production what comes to mind when the word "maintenance" is mentioned, and the most common responses he received were (i) high cost (ii) under utilised (iii) not highly valued (iv) bottom of the management group (v) non exploited opportunity for competitive advantage. All the responses were very negative in nature. According to Thomas [[1]], the feedback received by Palarchio mirrored his own experience as a maintenance consultant of 30 years. Following a series of interviews with people on what they do as engineers, Trevelyan [[3]] reported that "Maintenance is a major area of engineering in its own right and it was interesting that in our early interviews it hardly rated a mention".

The first phase of this research project will try to determine whether maintenance does have negative connotations. If so, the reasons for this will be investigated further in the second phase of the project. A possible explanation to investigate further could be that "maintenance" is seen as the culprit in major incidents. We know that some major accidents such as Chernobyl and Piper Alpha could be linked to in some way to maintenance and the procedures followed during maintenance. However, the general perception is often that "maintenance" – and not incorrect procedures during maintenance – caused the incidents. In Chernobyl for example, it was human errors made during preventive maintenance intervention that were the direct cause of the accident. On Piper Alpha, a series of procedural oversights and accidents led to the catastrophic destruction of the offshore gas platform and the death of people. The main oversight was not following the required procedures of informing the supervisors of the removal of a safety-critical component during a preventive maintenance action.

Another possibility to potentially investigate in the second phase of the project relates to the way in which accidents are covered by the news media. The public is sometimes subjected to repeated, visually impressive footage, supported by speculations on the cause of the accidents which are presented as facts. To the uninformed, "maintenance" can then become the easy explanation for the accident, instead of the errors made during maintenance.

4.2 Hypothesis 2 related to "People"

The strategic importance of maintenance is not fully appreciated in companies – especially not by strategic decision makers

The authors are aware of recent directives from top management of a large mining company to "*reduce maintenance cost by 30 million AUD over the next three years*". Numerous questions come to mind from such a directive: What would be considered as "maintenance costs savings"? Is maintenance cost seen in context by strategic decision makers? How sustainable will the cost reduction efforts be in the longer term? Is cost of "production loss due to failures" taken into account? Does top management fully appreciate the complex and interactive nature of maintenance?

Below are some thoughts on two of these questions.

"Is maintenance cost seen in context by strategic decision makers?" According to Hägerby and Johansson [[2]] the answer is "no"; and that the "true" cost of maintenance is neither fully understood nor presented in context in companies. Their results showed that for the seven Scandinavian companies reviewed, direct maintenance cost (e.g. cost of spares and labour) was overshadowed by the loss of production cost due to unavailability of the plant. The loss of production cost was up to 40% of company turnover, compared to the direct maintenance cost of approximately 8% of company turnover. Reducing direct maintenance cost in the short term might make some managers happy, but the potential effect could be an increase in failures in the longer term, with a resulting increase in production loss cost.

"Is maintenance well presented at board level or in top management?" According to Thomas [[1]] it might be true on paper that maintenance is presented on high levels in organisations, but in real life that is not true: "*Many organigrams show maintenance at the same level as production, marketing, finances, etc., but in my experience the reality is that maintenance is not part of "The Management Club*". According to O'Malley [[8]], understanding the importance of maintenance is not only a problem on higher management levels of organisations. Even the members in the maintenance team, such as reliability engineers, sometimes struggle to see the importance of maintenance - making it difficult for them to "sell" maintenance as a benefit to top management.

If maintenance is not well presented on the level where strategic decisions are made, and if the information provided to these strategic decision makers does not reflect the real benefits of maintenance, there is the risk that maintenance cost will be seen as an unnecessary expenditure - especially as it goes against human nature to spend time and money on something that has not happened (which is the basic principle of preventive maintenance). These decision makers might then put continuous and unrealistic pressure on the maintenance department to reduce the maintenance costs.

In the first phase of the research project, information will be gathered to determine how well maintenance is presented or understood by strategic decision makers. If there are indications that the maintenance is poorly presented, the second phase will try to determine the effects of the "poor" presentation of maintenance. Are there examples of where pressures on the maintenance team were considered "unrealistic", and if so has this pressure resulted in unsustainable cost cutting actions? Were the maintenance team members aware before implementing the actions that they were unsustainable?

4.3 Hypothesis 3 related to "People"

Maintenance management roles have a high turn-over, restricting the effectiveness of maintenance optimisation projects

Based on the authors' experience, people in maintenance management roles seem to stay in that specific role for approximately 2 years before moving on – either to a different role in the company, or to a different company. During consulting involvement over a period of six years, for example, the first author had to deal with three successive maintenance managers in one company. The different managers did not seem to fully appreciate previous maintenance optimisation efforts, and new strategies and changes were introduced by each, shortly after taking over the role. Besides duplication of efforts and costs, people affected by the changes also seemed to become "change-weary" and reluctant to embrace the changes.

Hayward [[9]] noted that most of the employees in the maintenance department of an American energy provider have been through some form of a change process between 1990 and 2000. They have been exposed to re-organisations, process re-engineering, competency modelling, continuous improvement, organisation development intervention and RCM. Few questioned the intent of these efforts, but many however questioned whether these changes were truly successful as no feedback was provided regarding the success rate.

The first phase of the research project will try to determine the turn-over of individuals in maintenance management roles, and how other people (e.g. the maintainers) experienced the changes in management. We also need to determine whether the turnover of maintenance managers differ from similar non-maintenance management levels. If is true that maintenance management roles have a high turnover, the second phase will gather more information on the effects of the high turn-over of managers. Were maintenance personnel for example subjected to continuous changes in approaches and strategies? Do these people show resistance towards new ideas because of this? Is there an underlying thought that a new idea will be replaced by another before long? Can evidence be found that organisations where maintenance managers stayed on for longer periods are more efficient and effective than other organisations where there had been a high turnover of maintenance managers?

4.4 Hypothesis 4 related to "People"

New maintenance managers and supervisors from different backgrounds experience different problems in their roles

The background of maintenance managers and supervisors seem to differ vastly. Some managers for example have engineering degrees while others have worked their way up through the ranks. No evidence could be found in literature that managers with different backgrounds have different approaches and success rates towards maintenance management. There was a hint, though, from Hägerby and Johansson [[2]] that some of the problems experienced by maintenance managers relate to their backgrounds. Some of the top performance workers in the maintenance field that were rewarded through internal promotions to maintenance supervisors and maintenance engineers, ended up managing their friends with whom they have worked for many years. The new managers experienced a lack of professional respect, and they struggled to "control" their workers. Thomas [[1]] is of the opinion that maintenance supervisors and team leaders lack management skills in general, and therefore are working at less than 60% of their potential. According to Orr [[10]] many technicians try the job of maintenance manager, but the pressures are such that few last.

In the first phase of the project, information will be gathered on the background of maintenance managers and supervisors. Managers will be asked how well they think their backgrounds prepared them for their new roles, and how well they were supported in their new roles. Finally they will be asked to comment on how they felt about their own performance – especially initially. If there are indications that managers from different backgrounds experience different problems, this will be investigated further in phase 2.

4.5 Hypothesis 5 related to "People"

Maintenance workforce is ageing, and there is no special preparation to transfer / extract the experience and knowledge from the older work force

According to Hayward [[9]], 40% of craft work force in the Department of Energy in USA will retire between 2002 and 2007, with the average age of the maintenance work force at the beginning of that period 50 yrs old. No statistics could be found on the average age of the maintenance work force in Australia, but anecdotal evidence suggests that this component of the work force in Australia is also ageing.

In the first phase of the project, the average age of the Australian maintenance work force will be determined. If there is evidence that the maintenance work force is ageing, the second phase will determine whether there is evidence of special preparations by companies to transfer the knowledge from an experienced work force to the younger people. One way of transferring this knowledge could be to get the older workforce to review and update all the job procedures attached to preventive maintenance tasks. These procedures often have mistakes, as highlighted by an engineer involved in the development of job procedures [[3]] "*To a certain extent, the engineer also relies on trained workers to detect and avoid mistakes or simple omissions in the instructions*". Another mechanism of transferring knowledge could be to provide inexperienced younger maintainers with experienced mentors for a period.

4.6 Hypothesis 6 related to "People"

Maintenance managers experience e-mail correspondence as excessive – especially in a corrective "fire-fighting" environment

One of the roles of the maintenance manager is normally to provide short term maintenance engineering support – especially where the repair of equipment is complex or non-standard, or where the correct functioning of the equipment is critical for the safe operation of the plant. Reliability engineers, on the other hand, are involved in the development and implementation of long term maintenance strategies.

From the authors' experience, the maintenance engineers are the first to suffer in a corrective "fire-fighting" maintenance environment - especially with a growing trend to communicate via elaborate e-mails. Between managing the technical aspects of high-priority and complex repairs, and processing hundreds of e-mails (of which a guesstimated 80% were sent to the maintenance manager "for information only") there seem to be little time left for the maintenance managers to reflect on the situation, and to introduce actions to move away from this corrective "fire-fighting" mode. Some maintenance managers expressed their reluctance to go on holiday or training courses, because they receive up to 250 e-mails *per day*! The effect of a holiday is quickly cancelled out by the daunting prospects of working through all those e-mails, which were not handled by the acting manager, during the manager's absence. In a study by Mintzberg [[11]] on what managers do, he noted that most managers that were required to read a lot of correspondence expressed a dislike for long memos, and that most long reports and correspondence were only scanned quickly. If maintenance managers also follow this approach of skimming through e-mails, important information could be missed or misunderstood, which can lead to dangerous or costly decisions.

In the initial phase of the project, it must be determined whether maintenance managers experience e-mail correspondence as excessive, and whether the capability to do their jobs is restricted by the amount of e-mail they have to deal with. If this is a valid hypothesis, the second phase of the project will determine (i) the quantity and typical length of e-mails received by maintenance managers; (ii) How much time the managers spend on processing e-mails; (iii) The percentage of e-mails sent to the managers "for information only", and how many e-mails require a reply from the manager; and (iv) Whether there is a link between the corrective maintenance and the amount of correspondence on the matter.

4.7 Hypothesis 1 related to "Processes and Strategies"

Tools and techniques used in maintenance optimisation projects do not always produce the expected results

The size of the maintenance industry resulted in the development of abundant tools and techniques to help the reliability and maintenance engineers in their efforts to optimise the maintenance. Technology in this field also changes extremely fast; and new terminology, tools and techniques are introduced continuously. Levitt [[6]] writes "*Every year some smart scientist, engineer, or maintenance professional comes up with one or two more techniques for predictive maintenance"*. On only one website (*http://www.plant-maintenance.com*) over 700 software packages managing some aspect of maintenance are advertised. There are no definite rules as to which tools and techniques are the best for a particular application, and a lot therefore relies on the experience (and biases) of reliability engineers, as it is expected of them to select the appropriate tools and techniques (O'Malley [[8]]) – a daunting task given all the choices, and easy to get it wrong.

Anecdotal evidence suggests that organisations are realising more and more that the tools and techniques used by them to optimise their maintenance do not always deliver what they expected. Leonard [[12]] speculates that many RCM initiatives in the industry failed because of the application of the cumbersome and inflexible RCM methodology in industrial contexts that differ widely from the aviation industry where the RCM technique was first applied. Feedback from individuals interviewed by Hägerby and Johansson [2] suggested that "first line maintenance" as a maintenance technique has been applied with limited success since the limited first line maintenance tasks expected from operators (such as lubrication) were not performed well. The majority of the maintenance people interviewed by Hägerby and Johansson [2] reported general dissatisfaction with the amount and type of maintenance done on their assets. Thomas [[1]] writes "*I cannot remember the number of times when visiting maintenance offices that I have seen expensive condition monitoring equipment gathering dust [……]. No doubt these were bought in good faith but they failed due to the lack of basic maintenance principles.*"

Information is required on (i) What techniques and/or tools are used by the reliability engineers to determine the "best" amount and type of maintenance (ii) On what did reliability engineers base their particular choices (iii) Is there evidence that these tools/techniques produce the expected results, and (iv) Are the facilitators that use the tools and/or techniques trained and experienced in using it.

4.8 Hypothesis 2 related to "Processes and Strategies"

Improper work flow processes and work order execution rules limit the effectiveness of maintenance optimisation projects

The maintenance work flow processes define the activities involved from identifying a requirement for maintenance to close-out of the work order or job card. The following are typical tasks that appear in maintenance work flow processes (i) create a request for maintenance work, (ii) review request for maintenance, (iii) determine the priority of the maintenance request and create a work order, (iv) scope the work, (v) approve the work order, (vi) execute the work, and (vi) capture maintenance history in the CMMS and close-out the work order/job card.

If the work flow process is to be used as the main communication tool between the various maintenance roles, the tasks listed above can be supported by additional information such as who should execute the task, and how should the task be done in the CMMS.

No researched evidence could yet be found on the importance of work flow processes in the management of maintenance, but the experience of the authors supports the value of a well-defined work flow process. One of the production plants where the work flow processes were updated by the first author reported a rise of 40% in the number of safety-critical preventive maintenance tasks completed in the agreed time. This dramatic increase was experienced only two months after introducing the new processes and re-training of the end-users on the changes and how to do their jobs in the CMMS. Hägerby and Johansson [[2]] noted that if the work flow and work prioritisation are not documented and enforced, people tend to do the interesting work orders first - leaving the boring work such as regular oil sampling for the next shift. This results in a backlog of "uninteresting", but essential maintenance work.

The first phase of the project will gather information to determine the value of work flow processes. Maintenance areas where the work flow processes are well defined and enforced will be compared with maintenance areas where the work flow processes are vague or not documented at all. Interrogating the CMMS as well as reviewing relevant documentation will be used to support and correlate the feedback from the interviewees. If the work flow processes do seem to play a substantial role in the effectiveness of maintenance optimisation projects, and if there is evidence that many maintenance departments underestimate the importance of work flow processes, the second phase will gather more detailed information.

4.9 Hypothesis 3 related to "Processes and Strategies"

The quality of the maintenance history data collected in the CMMS restricts useability of the data by reliability engineers

Yang & Strong [[14]] define data quality as *data fit for purpose,* and may be measured using key parameters that are important to the users of the data. What would be considered "poor data quality" to one user, might be seen as sufficient by another user. At the very least though, organisations would be expected to distinguish between the mandatory and optional information to be gathered in the CMMS.

From the authors' experience, many organisations are not serious about the quality of the maintenance history gathered in the CMMS. It is not clear to the maintainers (the main source of maintenance history) what information is mandatory to provide, data collected in the CMMS is not audited regularly, and feedback is not provided to the maintainers on how the data was used and what improvements resulted from the analyses of the data. Levitt [[6]] confirms this experience by reporting that "*Management has been wishy washy about requiring work orders and other maintenance record keeping to be accurate and complete*". An engineer interviewed by Trevelyan [[3]] remarked "*Often I find that we are collecting data but not using it effectively. Of course there are sometimes issues with the quality of the data and that needs to be taken into account before making too much use of it*".

The first phase of the project will determine (i) Whether information on the organisations' requirements with respect to maintenance data quality is readily available; (ii) Whether people are clear on what data fields in the CMMS need to be completed by whom, and when; (iii) Is there evidence that data is evaluated regularly to determine the accuracy and useability of the data? (iv) Is there evidence that the data is used by reliability engineers? and (v) Are the providers of the data informed on how the data is used?

4.10 Hypothesis 1 related to "CMMS"

CMMS implementations by people with limited understanding of maintenance and the organisation's business requirements result in CMMS considered ineffective and inefficient

Anecdotal evidence suggests that where "quasi-experts" (i.e. experts in the Computerised Maintenance Management Systems, but with little understanding of maintenance) were used to configure and later support the CMMS, people in the maintenance environment experience many frustrations in the day-to-day usage of the CMMS. Some individuals interviewed by Hägerby and Johansson [2] reported that "*The initial (PM) program was developed by people with little or no experience in operative maintenance. For this reason, unrealistic assumptions have been made, resulting in a PM program that contains* *several unnecessary or even totally inappropriate activities*". Mather [[15]] considers maintenance an area neither easily nor rapidly understood by those outside the discipline, and advises that only people with sufficient knowledge in both maintenance and the company's specific business requirements make critical decisions regarding the way in which the system should be configured. Levit [[6]] suggests using CMMS representatives as advisors only.

Information gathered in the first phase of this research project will determine how successful users of the CMMS considered the initial implementation of the CMMS, and to what they attribute the successes or failures. The major frustrations experienced by users of the CMMS also need to be identified. The replies will be evaluated before deciding on the need to add this hypothesis to the list to be further investigated in the second phase of the research project.

4.11 Hypothesis 2 related to "CMMS"

Insufficient maintenance of CMMS master data post go-live result in a CMMS not trusted by maintenance personnel

The "go-live" date of a new CMMS is normally broadcast widely in an organisation, and awaited with excitement. After the CMMS has gone operational though, it seems as if people are expected to get on with the "real job" of maintenance, and many organisations seem to limit the time and resources available to review and update the master data in the CMMS (data in the CMMS that do not change regularly as part of the day-to-day operations of the asset). The authors have noted that the maintenance data administrators for example are often of the lowest paid jobs in the maintenance environment. People interviewed by Hägerby and Johansson [2] reported that "*CMMS is used extensively, but these plans are not kept up to date and therefore not trusted by the maintenance personnel*."

During the first phase of the project, people will be asked whether they experienced problems related to the validity of the CMMS master data. They will also be asked whether it was easy to request updates in master data and whether they considered the approval of data changes prior to implementation sufficient. Feedback will determine whether this hypothesis will be addressed in more detail in the second phase.

5 CONCLUSION

Maintenance is a huge and fast changing industry. The main role players in this industry are the organisations with the maintenance needs (and they can spend up to half of their turnover on maintenance) and the providers of software, instrumentation, consulting, training etc. aimed at reducing the organisations' direct and indirect maintenance costs.

Despite the huge amounts spent by companies on maintenance, and despite the availability of tools, techniques, instrumentation, consultants, etc. to support the maintenance optimisation efforts, anecdotal evidence suggests that there are still many problems in maintenance. No structured research could be found on whether the main problems relate to issues with the people involved in the maintenance cycle, whether the issues relate to the strategies and processes applied in maintenance management or whether the majority of the issues are linked to the implementation and support of the CMMS.

This research will attempt to gather information in a structured and unbiased way, analyse the data using techniques that have proven their value in the social sciences research, with the aim of providing a better understanding of the major factors that have influenced (or still are influencing) the effectiveness of maintenance management efforts. This will provide a useful foundation for future research in this field.

6 REFERENCES

[1] Thomas C. (2005) Maintenance – A Business Centre Approach, retrieved from http://www.idcon.com

[2] Hägerby M. & Johansson M. (2003) Maintenance Performance Assessment – Strategies and Indicators, Master's thesis, Det Norske Veritas and Department of Production Economics, ISSN LiTH-IPE-Ex arb 2002:635

[3] Trevelyan J. (2005) Definition of Professional Engineering Roles - Revision 3, University of Western Australia.

[4] Moubray J. (1991) Reliability-centred Maintenance II, Oxford: Butterworth-Heinemann Ltd (ISBN 0 7506 0230 9).

[5] Mobley R.K. (1999) Root cause failure analysis, Newnes Publications, Butterworth-Heinemann Ltd, (ISBN 0-7506-7158-0) (http://www.newnespress.com).

[6] Levitt J. (2003) Complete guide to Preventive and Predictive Maintenance, New York, NY: Industrial Press Inc (ISBN 0-8311-3154-3).

[7] Palarchio G. (2004) The Physical Asset Management Profession in 2010, retrieved from www.mt-online.com/articles.

[8] O'Malley A. & Karyagina M. (2003) The Role of a Reliability Engineer, Worley Pty Ltd, Conference proceeding of ICOMS 2003.

[9] Hayward G.B. (2002) Developing a competency model based Maintenance Manager Qualification Program utilizing concept mapping methodology, Dissertation for Degree Doctor of Philosophy, Capella University, USA

[10] Orr J.E. (1996) Talking about machines: an Ethnography of a Modern Job, Ithaca, New York, Cornell University Press.

[11] Mintzberg H. (1973) The nature of managerial work, Harper and Row Publishers.

[12] Leonard J. (2004) Optimizing the role of the Maintenance Department, Maintenance Journal, October 2004 - retrieved from http://www.maintenancejournal.com.

[13] Weber A. (2004) So you think you have the right Maintenance Program, Maintenance Technology Journal retrieved from http://www.mt-online.com/articles/0705viewpoint.cfm.

[14] Yang W.L. & Strong D.M. (1996) Knowing-Why about data processes and Data Quality, Journal of Maintenance Information Systems, Winter 2003-4, Vol.20. No 3, pp 13-39

[15] Mather D. (2004) The Strategic Importance of Asset Management, Maintenance Journal, Issue October 2004 retrieved from http://www.maintenancejournal.com.

[16] Miller GA. (1956) The Magical Number Seven, Plus or Minus Two: Some limits on Our Capacity for Processing Information. The Psychological Review, vol. 63, pp.81-97 (copy available on http://www.well.com/user/smalin/miller.html)

OPTIMISING PLANT PERFORMANCE: A SYSTEMS THINKING APPROACH TO PROBLEM SOLVING

A Agresti [a]

[a] *Principal, Asset Capability Management Pty Ltd, www.assetcapability.com.au*

Abstract: Maintenance and Reliability Managers are responsible for solving plant performance problems that relate to maintenance activity or improvement activity within the operating environment. Knowingly or not, they manage maintenance activity using a series of interrelated processes; a system. Management of improvement activity should also use a system as improvement activity competes for resources and priority and adds a different complexity as it brings change; change to both behaviours and practices. Recognising that we operate within systems is fundamental to solving plant performance problems and remaining competitive. Activity within systems requires a different type of thinking to be fully effective. Thinking that fully explores the complexity of the cause and effect relationship between the constituent processes and the behaviours of the people that enact them. Systems Thinking is the key capability required by teams for describing and understanding problems to optimise plant performance. Subsequently, Maintenance and Reliability Managers that are unable to develop an understanding of Systems Thinking and apply it to establish an improvement management system will not realise the full performance potential of their plant.

Key Words: maintenance systems, performance improvement, problem solving, systems thinking

1 INTRODUCTION

Today's operating environment is changing at an increasing pace because technology, information, and competence are becoming business commodities [1]. Any organisation can purchase the latest technology, gain access to the same information, and have their people trained in the latest techniques. One way to remain competitive is for an operation to have the capability to improve plant performance through defect elimination.

The essential purpose of the maintenance function is the removal and elimination of defects to efficiently achieve stable plant performance according to plant availability and production requirements. Defect removal can be managed using either a reactive or planned maintenance approach. In a reactive approach there is an absence of any structured system; defects are dealt with as and when they are recognised. By contrast a planned approach uses a rigorous maintenance work management system (MWMS) which is far more effective as it ensures that resources are used appropriately, tasks are efficiently coordinated and roles are aligned to repeatable processes with managed outcomes such as availability and reliability.

Defects however are much more than technical plant performance problems. They can be caused by any number of interrelated technical, systemic or behavioural events which may happen in isolation or in any combination. To fully eliminate defects at their root cause first requires a comprehensive understanding of the interplays at work.

Systems Thinking is a practice that involves holistic consideration of the interplay between variables that influence a system; the discipline to recognise inherent patterns and related outcomes. It contains distinct principles that can highlight unique problems within an operational matrix of technology, people and work practices (processes). Systems Thinking is a key discipline essential for managing plant performance improvement. By correctly applying Systems Thinking an operation can establish the capability it needs for defect elimination to deliver optimal performance.

2 THE PATH TO SUSTAINABLE DEFECT ELIMINATION

Despite overwhelming proof that our lives and the operational environment are drastically changing, we continue to rely on the application of technically focused problem solving concepts and methodologies for improving plant performance. Maintenance and Reliability Managers, delegated responsibility for realising optimal plant performance, are trained to manage tangibles; equipment and plant performance by review of facts and data using argument and logical conclusion. They are not equipped to deal with the intangibles of individual perspective, knowledge, emotion, time and relationships.

The manufacturing and industrial operations of the future will need to effectively manage knowledge, emotion, time and relationships to realise the levels of productivity required to remain globally competitive. According to Senge [2], we are limited in dealing with such issues by our collective learning capability. The source of the problems that lead to the repeated inability to sustain significant improvement and realise optimal performance can not be solved by more expert advice, better consultants, or more committed managers. The source lies in our most basic ways of thinking; if this does not change, any new ‘input’ will end up producing the same fundamental unproductive types of actions.

Edward de Bono [3] also states that “The difference between brilliant and mediocre thinking lies not so much in our mental equipment as in how well we use it.” To be effective then, problem solving needs to go beyond realising technical outcomes to provide learning outcomes and we need a new way of thinking that incorporates the intangibles.

Systems Thinking provides the means to establish a holistic perspective of the issues limiting plant performance in an industrial operation. It provides learning outcomes that influence existing attitudes and practices within an organisation by allowing underlying norms, policies and objectives to be assessed as well as focusing on specific technical actions and results.

Application of Systems Thinking provides opportunities for collective inquiry so that it is possible for a team to develop new and better understanding of what to do, and how or when to do it. It provides the means to create enduring solutions to recurring or stubborn plant performance problems because performance improvement teams can work on the entire “problem system” rather than on specific elements or symptoms.

By developing an understanding of Systems Thinking, how it can be successfully applied, and its significance to realising sustainable improvement, Maintenance and Reliability Managers will have the means to manage these “intangibles” and realise optimal plant performance more efficiently and effectively than is otherwise possible.

3 AN OUTLINE OF SYSTEMS THINKING

The principles of Systems Thinking are derived from the field of System Dynamics founded by Professor Jay W. Forrester at the Massachusetts Institute of Technology during his work in electrical engineering in the 1950’s. System Thinking is a methodology for studying and managing complex feedback systems such as those found in business and other social systems. System Thinking focuses on feedback flows that occur throughout the parts of a system and the resulting system behaviours that are caused by those flows. Before exploring this type of thinking, let’s consider what is a system.

A system is a combination of interacting, interrelated, and interdependent elements that form a complex and unified whole. There are many familiar examples of systems such as the maintenance (defect removal) function in an operation, the respiratory system of the body, the relationships between predator and prey in nature, and the control system in an item of equipment. Specific characteristics are used to define a system:

Every system has a purpose within a larger system. (The purpose of the MWMS in an operational system is to provide an optimal level of plant availability and performance.)

- All system elements must be present for optimal performance of the system purpose. (The MWMS consists of people, equipment, and processes. If any one of these components are missing the system can not fully function.)
- System elements must be in a specific configuration to realise the system purpose. (Planning must occur before scheduling in the MWMS for work to be effectively executed)
- Systems change (behave differently) in response to feedback. (Feedback is information transmitted and returned through a possible chain of causes and effects from A to B and in turn from B to A that ultimately influences subsequent actions at A. The link between A and B can not be studied independently of the link between B and A to be able to predict how the system will behave. Correct results can only be obtained by a study of the whole system as a feedback system. For example, while snow skiing you negotiate a bend and turn too sharply. You receive visual signals (you see the trees ahead rushing toward you) that tell you that you are turning too hard. These signals form the feedback that stimulate you to change what you're doing (quickly shift your weight to the other leg) so you can get your self back on course.)
- Systems stability is maintained by making adjustments based on feedback received. (The average body temperature is generally around 37 degrees Celsius. If you get too hot, your body produces sweat to cool you down.

Systems Thinking is analogous to having a new set of eyes with which to look at your world to interpret outcomes and predict behaviours. By applying Systems Thinking you are able to distinguish between events, patterns and systems to identify the most appropriate action to take to manage a specific outcome or behaviour. For example suppose you suffer a bearing failure that causes an uncontrolled plant stoppage (a breakdown). This is an event. If you simply replace the bearing and get the plant running again you are reacting. (Nothing has been done to prevent the breakdown recurring.) If you decide to repair the equipment and study the failed bearing you might notice that the bearings in this equipment seem to fail more than in other equipment; you would be recognising a pattern.

If you decide to inspect the bearing more often you are adapting. (Nothing has been done yet to prevent the breakdown recurrence.) If you consider the systems that influence the pattern of bearing failures (e.g., operating practice, vibration levels and the lubricant used) you can decide to conduct vibration monitoring and establishing standards for lubricants and operating practices to introduce (behavioural) change. This is positive action to prevent breakdown recurrence.

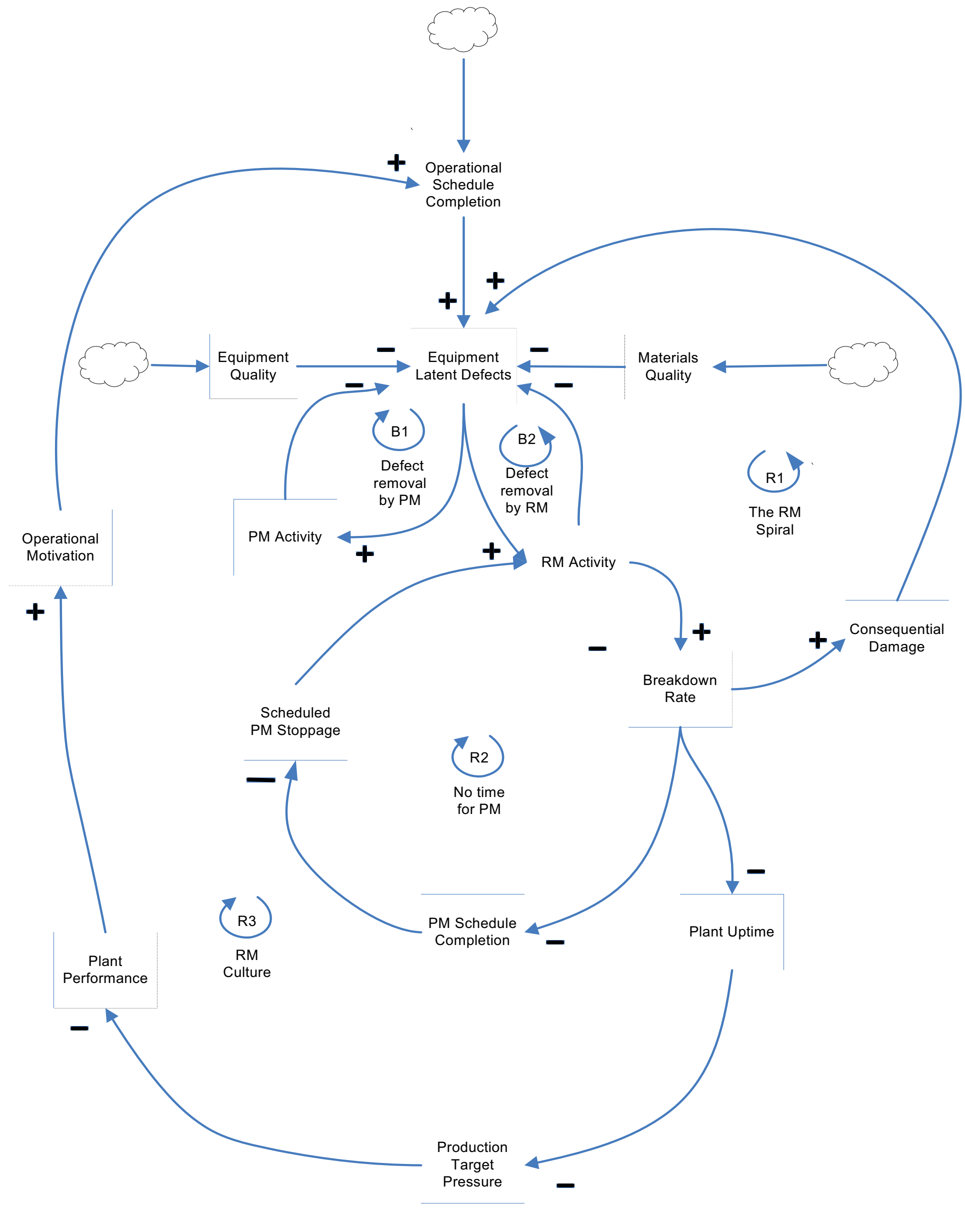


Figure 1 Causal Loop diagram for Defect Removal by RM

Systems Thinking uses casual loop diagrams, like that shown in Figure 1, to graphically depict the range of relationships that exist between variables that form the system under consideration and to develop an understanding of consequential system behaviours. It applies a nomenclature of stocks (variables that change over time) and flows (arrows that indicate how information flows between variables to affect each other) to describe the connections between the various feedback loops within a system and specific terminology to describe system behaviours. This includes the use of cloud symbols to indicate system boundaries.

There are two types of processes used to construct causal loop diagrams like that shown in Figure 1. Reinforcing processes, denoted by an "R", where feedback flows generate exponential growth or collapse and balancing processes, denoted by a "B", where feedback flows help a system maintain stability. Each type of process may include some form of delay, between a cause and an effect, as a regulator of systemic behaviour. The significance of delay is highlighted by considering the possible implications when a specific cause takes a long time to exert its effect, or when cause and effect are geographically separated..

There are many practical examples of reinforcing and balancing cycles all around and within us. The global manufacturing impact of China's industrialisation, the foreign currency trading that lead to massive corporate losses by NAB in 2004, and the potential for a pandemic when gene mutations in Avian flue resulted in the H5N1 virus are all examples of reinforcing cycles. The ability of the human body to maintain a temperature of 37 degrees Celsius, the stability that occurs in natural ecosystems, and the challenges faced by an organisation when trying to make an improvement by changing the way people work are all examples of balancing processes.

4 SYSTEMS THINKING AND DEFECT REMOVAL

A representation of the reinforcing and balancing loops found in the processes used to remove latent defects present in equipment is shown in Figure 1. The cloud symbols used in the diagram indicate that the defect removal system is part of a larger system that includes defect elimination. By way of example, consider what can happen if planned maintenance (PM) activity is reduced in an effort to cut costs and increase plant availability by reducing scheduled downtime:

- the level (stock) of equipment latent defects is primarily added to by operational activity (and potentially reduced by the quality of equipment and materials used)
- defects are removed according to the volume of either planned (PM) or reactive (RM) activity
- reduced PM activity means latent defects will accumulate to the point where equipment reliability (and plant performance) begins to decline without this being obvious (because of the delay between cause and effect) until frequency of breakdown events begins to increase which will increase consequential damage and initiate the RM spiral (R1)
- an increasing breakdown rate compounds the fact that scheduled PM work can not be completed because there is no time for PM, so PM backlog grows as schedule completion falls (R2)
- as plant uptime decreases, production pressure increases and plant performance decreases (operational and safety performance drops off and absenteeism climbs); there will be some delay but eventually people stop reporting identified defects and or providing the level of detail required to effectively plan defect removal work (operations and maintenance people become frustrated and loose motivation). This leads to lower effective completion of scheduled operating activity that results in increased introduction of latent defects. Since in this environment defects are mainly removed by RM activity, maintenance personnel are rewarded for efficient breakdown recovery in an increasing effort to achieve operational output targets (reinforcing a reactive pattern of behaviour: R3 – RM Culture)

If you don't notice this level of feedback for an extended period of time, you most certainly will not realise that the drop in plant performance, equipment reliability and performance of personnel are connected to the reduced plant downtime and cost reductions you introduced "ages ago." (It is quite possible that you might even panic about declining equipment reliability and plant performance and introduce more rigorous PM activity (which will not be able to be completed) to try to correct equipment reliability and improve operational performance!)

This example also shows how systems structure directly influences organisational behaviour. That is, system structure generates recurrent "signature" patterns of behaviours inherent within it. These recurrent patterns of behaviour within the organisation are those balancing processes that become entrenched and difficult to change.

The B2 balancing loop of "Defect removal by RM" is sustained by the R1 reinforcing loop of "The RM Spiral" and the R2 loop of "No time for PM" that further entrenches RM behaviour B2. These recurrent behaviours are represented as System Archetypes within the discipline of Systems Thinking.

If placed in this same reactive "system", regardless of their individuality, all people will tend to produce the same results because a system tends to cause its own behaviour. People within this defect removal system will all behave in a reactive manner because "that is the way we always work here", becomes the generally accepted norm. By recognising these patterns and acting to change behaviour by increasing the downtime rate and subsequently increase PM activity it is possible to shift the defect removal system back into stability that satisfies the maintenance purpose.

Most industrial or manufacturing operations will be using some form of a MWMS and be in either a reactive or planned operational domain. Most will attempt to improve plant performance; move from the reactive domain into the planned domain by introduction of the latest method or technology. This is good but it does nothing to achieve improved plant performance by defect elimination. The prime outcomes delivered from a focus on the planned domain are effectiveness and efficiency of defect removal activity. It does not address organisational and technical system dynamics essential for defect elimination. Eventually, performance will drop below the required plant performance baseline and some intervention will be needed again.

Ledet and Monus [4] give an example of why a focus on what they call "the planned domain" can produce good results but is not sufficient for realising sustainable and optimal plant performance. They explain how Dupont realised a 20% reduction in total annual maintenance cost, which gave a maintenance cost ratio of 2.2%, over a 10 year period across 50 operating plants. After 10 years performance began to deteriorate and 15 years later (the delay between cause and effect) the improvement gains had disappeared and they had to start again. By contrast, a comparison operation realised around an 80% reduction in repair activity and cost as well as improved reliability, production output, and safety performance. This performance was able to be sustained through numerous changes in ownership because the comparison operation "created" more time for their workforce to be proactive and focus on improvement (defect elimination) activity.

5 SYSTEMS THINKING AND DEFECT ELIMINATION

There are two types of systems; natural living systems and created nonliving systems. In a plant operation context these are the organisation's culture (people and attitudes; elements of the natural behavioural systems) and technical (plant and practices; elements of the created equipment and management systems). According to Senge [5], all growth (improvement) is the result of the interplay between growth (reinforcing) processes and limiting (balancing) processes that define a system. Realising optimal plant performance requires understanding of the reinforcing processes and what is needed to activate them, and addressing the balances that keep optimal performance from occurring within an organisation's behavioural and practice systems.

There are two types of leadership; Leadership of Operations and Leadership of Change. In a plant operation context these are the management of defect removal and the management of defect elimination. The former involves establishing and maintaining steady state plant performance using a MWMS. The latter involves development and introduction of change to realise plant performance improvement. Leadership requires guidance of team activity. Making sure that people know what to do and what is expected of them. It makes sense that leadership of change, guidance of teams to identify and implement improvements, requires an improvement management system.

It is a common error for an operation to try to improve plant performance by identifying best practice and then implementing a corporate improvement program. In most cases this does not work because this approach takes no account of the systems archetypes that exist. The problem is that nothing is done to address system archetypes, introduce behavioural change, and ensure the change process is completed. This happens because the effort essential for establishing an effective improvement management system (defect elimination capability) is not appropriately acknowledged and the effect of "systems delay" means that resultant adverse consequences are not appropriately recognised for some time, if at all. Building an improvement management system takes time; usually longer than the typical management position tenure.

Another trap organisations often fall in to is thinking that defect elimination is solely about analysis of plant data or application of panacea techniques such as Six Sigma, Lean, RCM, FMECA, or RCA. It is not. There is definitely a place for this type of activity but it must be considered in the greater context (system) of the operation. Providing consideration for the activities your entire operation performs and the time factors associated with those activities across both the production and maintenance functions. By doing this and applying Systems Thinking concepts an industrial operation will derive the key capability for realising optimal plant performance. Using Systems Thinking to establishing a system for managing improvement (defect elimination) Maintenance and Reliability Managers can ensure improvement initiatives focus on the result they are trying to produce and build a key capability necessary to sustain optimal performance.

6 AN IMPROVEMENT MANAGEMENT SYSTEMS MODEL

Having the capability to effectively conceive, prioritise, deliver and manage a plethora of competing improvement initiatives requires a system for managing improvement to realising optimal plant performance. A system founded on Systems Thinking concepts that make use of both technical (tools and techniques) and learning (new knowledge and attitude) inputs to derive desired plant performance outcomes. The improvement management system (IMS) model shown in Figure 2 is an example of such a system.

The IMS model contains three coordinated and linked levels of "leadership teams" to deliver effective plant performance improvement and realise the operational goals of the business. The lead and steering teams link to ensure all improvement activity is aligned and prioritised according to prevailing business operational requirements and to formulate business policy to ensure consistency and capability of delivery. The steering and project teams link to provide leadership of change and ensure effective resource application minimising chaos within existing processes or operational activity. All levels of people within the IMS are aligned to ensure shared understanding of the improvement management system purpose for realising optimal plant performance.

The IMS model provides a framework whereby the leadership teams share common corporate strategic goals and focus on such strategic objectives as cultural alignment, change management, and resource optimisation. Improvement projects are conceived, substantiated, justified, and prioritised according to an optimisation policy, in preparation for subsequent implementation. The IMS model promotes the use of facts and data to identify and eliminate chronic problems. It facilitates cross functional participation and shared responsibility between maintenance and operations for improvement projects to ensure collaborative learning takes place. This eliminates the risk of competition for priority with day to day functional work, and ensures available resources both across and within functions is effectively deployed. Application of the IMS model ensures a "systems" perspective to minimise risk of improving one aspect of plant performance to the detriment of the whole.

The IMS model describes the stocks and flows of information associated with improvement management. The feedback loops implicit in the model exist by virtue of routine performance monitoring. Performance monitoring of the IMS outcomes (operational performance), the IMS itself as well as measuring the achievement of project outcomes. Business goals are applied to set IMS goals and the system is managed to best achieve steady state performance towards achieving these goals. Improvement projects may be established by either strategic or operational improvement request. However, under steady state business and operational conditions the majority of improvement activity is the result of considering actual plant performance against required standards of performance using "history maps" and deviations are investigated using an RCA methodology.

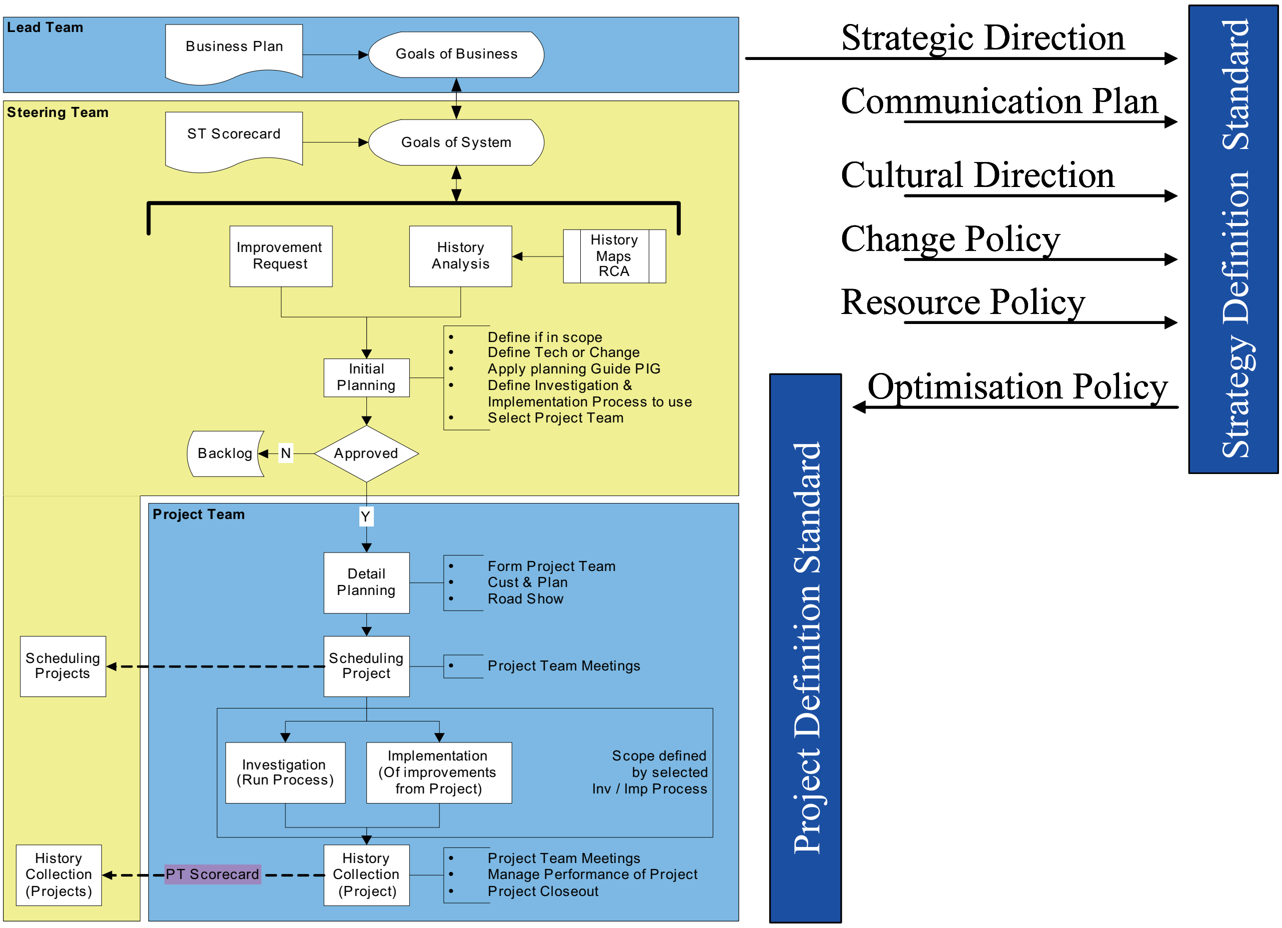


Figure 2 Improvement Management System Model

The IMS model incorporates planning processes to develop understanding of readiness for change and what things must be addressed to ensure performance does not regress. This is done by applying specific actions to:

1. Articulate the business need that will be addressed by making a change.
2. Empower the workforce to change the way work is done.
3. Deal with issues of authority and motivation.

These are addressed in the IMS model through appropriate, meaningful and planned communications, project score cards, investigation and implementation methodologies and selection and assembly of project teams to ensure optimum participation. Planning processes should include consideration of what people across the operation will need to learn to address both behavioural and technical system dynamics. This is made possible by use of a planning guide to define the “problem system” and help people logically identify their assumptions about structure, behaviour, and expected outcomes as part of an improvement initiative.

Additionally, planning considers the three prime dimensions that an operation must consider when determining how to optimise performance; time, quality, and cost. In doing so it provides guidance to select the method best suited to the business need and impact on profitability. Planning within the IMS model incorporates the approach espoused by Nave [6] to select between the high level techniques for improvement such as RCM, Six Sigma, Lean Manufacturing, and Theory of Constraints. All these techniques target improved business performance but each focus on different plant performance dimensions. Selection of the most appropriate technique is a critical step.

Project close out includes activity to ensure the change process is completed and can be locked in to routine practice using performance management to sustain the improvement.

7 CONCLUSION

Making improvement to deliver sustainable and optimal plant performance requires a fundamental shift in thinking. Firstly we need to recognise that there are separate and interlinking management systems required to manage both maintenance (defect removal) and improvement (defect elimination) and that improvement requires change to existing behaviours and practices and is ultimately about cultural change (change the way we do things).

Introducing an improvement management system will ensure best return on both operational and capital improvement resources and achieve sustained optimal plant performance. It will provide a stable and repeatable learning and knowledge management capability. This will provide mutual benefits to both Maintenance and Production managers' of less stress, more time to think, capability to jointly respond with considered haste to changes in operating demands, and high probability of consistently achieving plant safety, environmental, cost and output targets.

Secondly we need to understand the nature of growth processes (forces that aid our efforts) and how to activate them by adopting a systems thinking approach. The application of Systems Thinking will allow the cause and effect relationships within "problem systems" and the timing of subsequent system behaviour to be understood and provide a capability for optimising plant performance that is more than just problem solving. Systems Thinking is a fundamental capability for Maintenance and Reliability Managers to equip them to deal with both the tangibles and intangibles of plant performance. People start seeing and dealing with interdependencies and deeper causes of problems only as they develop the skills of Systems Thinking.

It is naïve and overly simplistic to assume that a change in one part of an operation will not impact in some way on another part. You are never acting alone; you are always part of a bigger system. Finding a solution to 'your' plant reliability or performance problem can affect the entire operation. Establishing an improvement management system developed from a solid basis in Systems Thinking will result in a profound change in the relationships and behaviour of individuals within the operation. Maintenance and production people will be happier, more motivated, and more committed because they will understand how to effectively collaborate to achieve a common goal of optimal plant performance. A goal that includes equal consideration of time, cost, safety (environmental and human) and quality as the key business drivers of operational performance.

Like Solomon said: nothing under the sun is new. The prime concepts and principles assembled in the IMS model, largely based on the published work of Peter Senge, have origins in the early 1930's. It is a matter of combining engineering and management discipline with behavioural science and that is how the IMS model has been developed from its origins in 1995. More than ten years later the IMS concept is still relatively undiscovered. The fact is managers usually prefer to follow rather than lead; it is less risky. Those Maintenance and Reliability managers that choose to be pioneering leaders and establish a capability for realising sustainable and optimal plant performance will provide a valuable solution in the globally competitive market.

8 REFERENCES

1. de Bono E. (2005) *The Six Value Medals.* London: Vermilion
2. Senge PM. (1990) The Fifth Discipline: The Art and Practice of the Learning Organization. New York, NY: Doubleday.
3. Lyons JC & de Bono E. (2004) *Marketing without money.* Niddrie, Vic: Pennon Publishing
4. Ledet W.P., Monus P. A. (2005) Why planned maintenance Is not enough anymore: A comparison of success at DuPont vs. BP Lima, *Reliability Magazine Vol. 10, Issue 6*. pp 23 – 27.
5. Senge PM., Kleiner, A., Roberts, C., Ross, R. Roth, G., & Smith, B., (2002) *The Dance of Change: The Challenges of Sustaining Momentum in Learning Organisations.* London, Eng: Nicholas Brealey
6. Nave D. (March 2002) How To Compare Six Sigma, Lean and the Theory of Constraints: *A framework for choosing what's best for your organization* Quality Progress pp 73 - 78

A MODEL TO DIAGNOSE THE DEVIATION IN THE MAINTENANCE PERFORMANCE MEASURES

Basim Al-Najjar

Department of Terotechnology, School of Technology and Design
Basim.Al-Najjar@vxu.se

Abstract: In order to achieve the strategic goals, the company should utilise an efficient data gathering and analysis system for mapping the technical and economic situation, following up the development and detecting technical and economic deviations at an early stage to plan and perform necessary maintenance actions on time. In many cases, measuring and following up the development of relevant and well selected technical and economic performance measures can be more applicable for mapping the situation and making cost-effective decisions compared with monitoring big mount of data. But, in many cases it demands a reliable technique for interpretation of the behaviour of these performance measures especially if they are not linearly interrelated and if some of them reflect a combination of technical and financial impact. The problem addressed in this paper is; how is it possible to enhance the ability of detecting significant deviations in the maintenance performance measures at an early stage and tracing their root causes by using data- and knowledge base and inference engine? The major result achieved in this paper is a model developed for interpreting the changes in the maintenance performance measures and tracing their basic causes, which is verified in an example with typical data. The model consists of five modules; what performance measures to chose, how and why an eventual deviation on the performance measures occur, what to do to eliminate and prevent their reoccurrence and mapping of the current situation technically and financially. The main conclusion is that applying the model would enhance the ability of detecting changes in the maintenance performance through identifying the causes for elimination. In Al-Najjar et al (2004), a model for how to identify the measurable variables, which are needed to develop measures for monitoring maintenance performance behaviour systematically, is introduced and discussed. Also, five maintenance performance measures are proposed and applied. Furthermore, an additional model manual analysis of the trend of maintenance performance measures, for an overall assessment of the company's situation, is presented. In this paper, the idea of manual analysis presented in Al-Najjar et al (2004) is systemised.

Key Words: Maintenance performance measures, knowledge base, inference engine, performance measure interpretation

1 INTRODUCTION

In many cases, it is important to apply relevant indicators for monitoring and following up the development of the performance of maintenance and production processes. These indicators can be simple, such as the number of items produced and number of failures, or complex, such as the cost of producing one high quality item and maintenance cost paid for every produced high quality product.

Many authors such as Kennerley and Neely (2003), Al-Najjar et al (2004), highlight the importance of properly identified the performance measures required for following up the work done to achieve company's strategic objectives and survive daily competition. According to Kaplan and Norton (2004), 70 to 90 percent of the companies fail to realize success from their strategies due to lack of relevant data. For companies to be successful and able to accomplish their strategies, it is necessary to implement and control their performance measures in all working levels, such as strategic, operative and tactical daily work, Waeyenberg and Pintelon (2002), Kaplan and Norton (2004) and Al-Najjar et al (2004). Nonaka (1991) argues that it is not just the connection between performance measures and strategy that is important, but also the knowledge required for the organization to achieve their strategic goals. Since this is important there is a need to consistently develop new experience and knowledge in different forma, such as new tools, methods and technologies.

Turban and Aronson (1997), Pintelon and Van Puyvdelde (1997) and Sherwin (2000) argue for the need of investments to secure successful implementation of performance measures and strategies. In general, the major objectives of applying information systems are to gather, analysis and present information and knowledge. Integration of the performance measures with a knowledgebase and an inference engine can provide the manager the required information, knowledge and ability to monitor and interpret the performance measures for making cost effective decisions.

The problem addressed in this paper is; how is it possible to enhance the ability of detecting significant deviations in the maintenance performance measures at an early stage and tracing their root causes by using knowledge base and inference engine? The aim of this paper is to develop a model describing the possibility of using a data- and knowledge-base and inference engine for real-time diagnosis of the behaviour of the performance measures and effectively detecting technical and economic deviations from predetermined values at an early stage. In Section 2, interpretation of maintenance performance measures is discussed. A model describes the utilisation of the data- and knowledgebase together with inference engine for detecting, analysis performance measure deviations and tracing the reasons behind that is developed in Section 3. The model is applied in an illustrative example in Section 4, which is followed by conclusions.

2 INTERPRETATION OF MAINTENANCE PERFORMANCE MEASURES

In this paper the maintenance performance measure is defined as the indicator that provides the information required for mapping and judging the performance level of maintenance technically, financially or combined. Identifying the relevant data required for accurate monitoring of maintenance performance is discussed in Kans (2005) and Al-Najjar et al (2004). Applying few well-selected performance measures for monitoring maintenance makes the procedure of judging its performance and consequently decision-making process much easier compared with the case when using a big amount of data for the same purposes. But, in the same time it demands better techniques to interpret the behaviour of the performance measures and achieve accurate diagnosis of the root causes, Al-Najjar et al (2004). We assume that the measures required for monitoring the performance of maintenance are already selected or developed by applying special techniques, such as top-down analysis method suggested by Al-Najjar et al (2004), Al-Najjar and Kans (2005) and Kans (2005).

Using just one performance measure for detecting changes in any process, such as the case when controlling the quality of a product using specific dimension to judge whether it is accepted or rejected is easy to interpret. But, when the maintenance or production cost is monitored, the increase in any of these two measures may mean many things, such as real increase in the direct maintenance cost, raw material, energy or operating cost.

The increase in the maintenance cost may be beneficial. For example, a chef manager would welcome the increase in maintenance cost if its share in the cost of producing one-item of high quality is reduced, i.e. the number of high quality items has been increased due to more cost-effective maintenance. The same confusion can happen when the production cost increases. This increment is not necessary an undesirable result especially if it is due to the increase in the production time and/or speed and not because of higher maintenance cost per producing one item or due to production losses. The situation will be much more complicated if there exist many performance measures that should be considered for mapping, analysis and evaluation of the production and maintenance processes to make cost-effective decisions.

We can easily trace the root causes lying behind changes in the technical or financial performance measures, such as machine failures and availability, and maintenance cost and production costs, respectively. But, it will not be equally easy to race the root causes when monitoring combined measures mixes technical and financial terms such as direct maintenance cost divided by overall equipment effectiveness, or direct maintenance cost divided by the number of quality items (or tons produced). This is because the increment or reduction in the performance measure value may mean different based on how the measure is developed, i.e. how the information provided by the measure is aggregated. Therefore, interpretation of the changes in the values of the performance measures and tracing the root causes behind them can be a big task, which is not easy to perform manually for achieving cost-effective decisions. This is why systemising this analysis and diagnosing process is important for carrying out a real-time interpretation effectively.

In general, the main objectives of applying relevant performance measures is to detect deviations in the cost and condition of the production and maintenance processes for taking the actions required at an early stage with less labour and cost. Also, the analysis and diagnosis of the deviations in the performance measures will lead to better results if it is associated with identifying the root-causes lie behind changes. Further, recommending what to do for treating the deviations would help avoiding their re-occurrence.

In Al-Najjar et al (2004), a model for identify the variables that are required to develop measures for monitoring maintenance performance behaviour systematically, is introduced and discussed. Also, five maintenance performance measures are proposed and applied. Furthermore, a model for manual analysis of the trend of maintenance performance measures, for an overall assessment of the company's situation, is presented. In this paper, the idea of manual analysis presented in Al-Najjar et al (2004) is systemised.

3 MODEL DEVELOPMENT

In this section, we consider that the model is necessary to detect significant deviations in the maintenance performance when several measures are used for monitoring maintenance and production processes. In order to achieve reliable mapping, analysis and evaluation of the maintenance performance, and to recommend what to do when significant deviations are detected, the model is constructed to perform the following five tasks, i.e. to

1. Identify what to focus on out of all the performance measures being monitored, i.e. performance measures that undergoing significant changes. This step is necessary because, in many cases, there are many monitoring variable

(measures) to be considered. Therefore, identifying where to start the investigation and why it is usually not an easy task

2. Analyse the deviations to answer how it is happened, i.e. describing the mechanisms of the initiation and development of the changes
3. Identify the root causes behind these changes, which is required for the prevention of reoccurrence of the same problems
4. Recommend what to do at an early stage to restore the process to as good as new (or as good as before), and finally
5. Assess the measures that are relevant for mapping and evaluating maintenance performance

These five phases of the model performance are carried out by equivalent number of modules. The first module is called “What to analyse”, which is responsible of, see also Fig. 1;

1. Arranging a list consists of the values of all the performance measures (x_1,..., x_n) that have been selected for performance monitoring
2. Establishing a list of the reference levels (y_1,.., y_n) for every performance measure introduced in 1 above. These reference levels are important for distinguishing the changes in the values of the performance measures and its severity. The number of the reference levels for each performance measure can be considered three; one at the normal level and one at the action level, i.e. an action should be done as soon as the measure value exceeds that level. The third level (warning level) can be positioned in between these two levels (normal and action levels) for indicating if the measure is undergoing a real change (but still acceptable) and not due to temporarily disturbances, Al-Najjar (1997)
3. The trends of these changes, i.e. changes in time of every performance measure, ease following up their developments. Association of these trends with the reference levels, such as normal, still acceptable/warning and action levels would provide the possibility of detecting significant deviations at an early stage.
4. The inference engine completes the work by reading whether any of these changes exceed the warning level and produce a prioritising list to prioritise the investigations

The second module, which is called How is responsible of investigating the root causes behind the deviations in the performance measures. To complete the work done in the What module, the investigation of the performance measures topping the prioritising list, i.e. those which undergoing significant changes, should be analysed by the inference engine in the How module. Previous technical experience, knowledge and analysis results of the possible reasons and the problem developing mechanisms that may cause deviations in the performance measures can be utilised when building the inference engine for identifying where and how to investigate. The reasons behind the changes in the prioritised measures can be several. Therefore, the inference engine investigates all these possible reasons individually and in combination with other reasons to detect possible interactions, see Fig. 1.

The third module (Why) is responsible of summarise the work done by How module by identifying the most probable reasons behind the deviations in the prioritised performance measures. The output of the Why module eases the task of recommending what to do, i.e. to suggest either to just handle the current situation, or to eliminate the roots causes or both, which is the task of the fourth module. The fifth module is responsible of mapping maintenance and production processes technically and financially, which is necessary for evaluating the maintenance performance.

The data- and knowledgebase that are involved in this model ought to include all the relevant data required by the inference engine to perform this task. It includes the knowledge and the previous experience concerning detecting, analysis and the results of handling previous problems. The data- and knowledgebase is necessary for improving the model continuously and cost-effectively. The model provides a possibility to assess not just technical or financial performance measures but also combined measures such as those expressed in the module performance measures assessment, i.e. the fifth module

4 APPLYING EXAMPLE

In this section, the model is described by an example treating one measure for monitoring the performance of maintenance through using the direct maintenance cost (DMC) divided by the high quality production (HQP) measured in tons or items, which is denoted by DMC/HQP. The performance measure DMC/HQP reflects the maintenance contribution in the production of each item or ton assuming that the quality control system is working properly. In other words, it means how much the company pays for maintaining the production system per each item or ton. Also, it describes the reliability of the production process.

When the list of the performance measures and trends of their behaviour are described, the inference engine will be able to prioritise the measures of the most critical values, i.e. the measures whose values exceeded a predetermined level. Let, for instance, one of these levels being the warning level. Assume that DMC/HQP exceeded the warning level. In general, the levels of the performance measures can be decided based on the company's policy.

When a measure exceeds a warning level it is always better to investigate the root causes and the developing mechanisms behind this significant change. Also, it is necessary to investigate interactions of these causes, if it possible, for identifying and cost-effectively eliminating them for restoring the process to its previous condition.

In this example, we considered the following factors are the most probable causes behind the increment of DMC/HQP, see also Fig.2:

1. Less accepted production; lower production quality due to inefficient or lack of maintenance.
2. Higher production rate; high production speed leads to faster deterioration in the machine condition, more stoppages, longer stoppage time and consequently more direct maintenance cost
3. Longer production time; it means also faster deterioration and shorter life length of the machinery, and consequently more direct maintenance cost
4. Quality of maintenance actions; it influences the maintenance impact on the machines and production process. The bad maintenance quality results in more frequent stoppages, less production and higher direct maintenance cost

In order to identify the basic causes behind maintenance performance deviation, the inference engine would analyse technical and financial past data of the machine and process. Identifying the interactions between the root causes those mentioned above are important to clarify the situation better. When the root causes are identified, the recommendation for what to do in order to eliminate the problem and prevent its reoccurrence is the next step to be accomplished by the inference engine.

Reporting what happened together with the assessment of the measures of the maintenance performance eases for decision maker to achieve a cost-effective maintenance action.

5 CONCLUSIONS

Monitoring and following up the development in production and maintenance processes automatically through applying special performance measures demands using special rules to implement a data- and knowledgebase and inference engine to accomplish a cost effective maintenance decision. Applying the model developed in Section 3 offers opportunity to diagnose and interpret the behaviour of the performance measures monitored. In general, the performance measures can be many in complex production processes. Using this model, detecting significant deviations in the performance measures can be done effectively does matter how many they are. Identification of the root causes behind these deviations enables the user to take the cost-effective actions required at an early stage.

It was obvious from the applied example how important are the past data from the working areas related to the performance measure in question. The applied example shows the possibility of monitoring even more complicated performance measures such as direct maintenance cost/high quality production, i.e. combined technical and financial measures, if the required data are available.

The knowledge available that can be utilised for identifying and analysis production and maintenance problems is enormous. This knowledge is not easy to utilise effectively without better data quality and coverage from the working areas relevant to the problems under consideration. A knowledge and data based analysis system using the technology of data- and knowledgebase and inference engine opens for the companies a possibility of applying real-time cost-effective maintenance decisions. At the same time it demands that the company should store specific data and knowledge required for applying and enhancing the model involved. This leads to decisions of less uncertainty, which in turn leads to lower level of risk and cost diminishing. Further, this also leads to more effective and efficient use of the companies' resources for both short and long term towards higher profitability and better competitiveness.

6 REFENCES

[1] Al-Najjar, B, Hanson, M.O, Sunnegårdh, P., (2004). Benchmarking of Maintenance Performance: A Case Study in two manufacturers of furniture

[2] Al-Najjar, B (1997), Condition Based Maintenance: Selection and Improvement of a cost-effective vibration based policy in rolling element bearings. Doctoral thesis, ISSN 0280-722X, ISNR LUTMDN/TMIO-1006-SE, ISBN 91-628-2545-X, Lund University, Inst of Industrial Engineering, Sweden.

[3] Al-Najjar, B. and Kans, M (2005). A Model to Identify Relevant Data for Accurate Problem Tracing and Localisation, and Cost-effective Decisions: A Case Study. To be published

[4] Kans, M. (2005). On the identification and utilisation of relevant data applying cost-effective maintenance. Report, no 14 (Tekn. Lic.), School of Technology and Design, Växjö University, Sweden

[5] Kaplan, R.S and Norton. D.P (2004), "Strategy Maps - converting intangible assets into tangible outcomes", Harvard Business School Press, Boston, Massachusetts, ISBN 1-50139-134-2, USA

[6] Kennerley M and Neely A (2003) "Measuring performance in a changing business environment", International journal of Operations & Production Management, Vol.23, No. 2, pp. 213-229

[7] Nonaka, I (1991) "Harvard Business Review on Knowledge management", Harvard Business School Press

[8] Pintelon, L,& Van Puyvdelde, F, (1997) " Maintenance Performance Reporting Systems: Some Experiences", Central for Industrial Management, Leuve, Belgium and Glaverbel, Mol, Belgium, Journal of Quality in Maintenance Engineering. Vol 3 No 1, 1997, pp, 4-15.

[9] Turban, E, & Aronson, J, (1997), "Decision Support Systems and Intelligent Systems", 5th edition, Prentice Hall PTR Upper Saddle River, NJ, USA.

[10] Sherwin, D (2000), "A Review of Overall Models for Maintenance Management", Journal of Quality in Management, Vol. 6 No. 3. 2000. pp. 138-164.

[11] Waeyenbergh, G, Pintelon, L (2002), "A Framework for Maintenance Concept Development", International Journal of Production Economics, pp.299-313

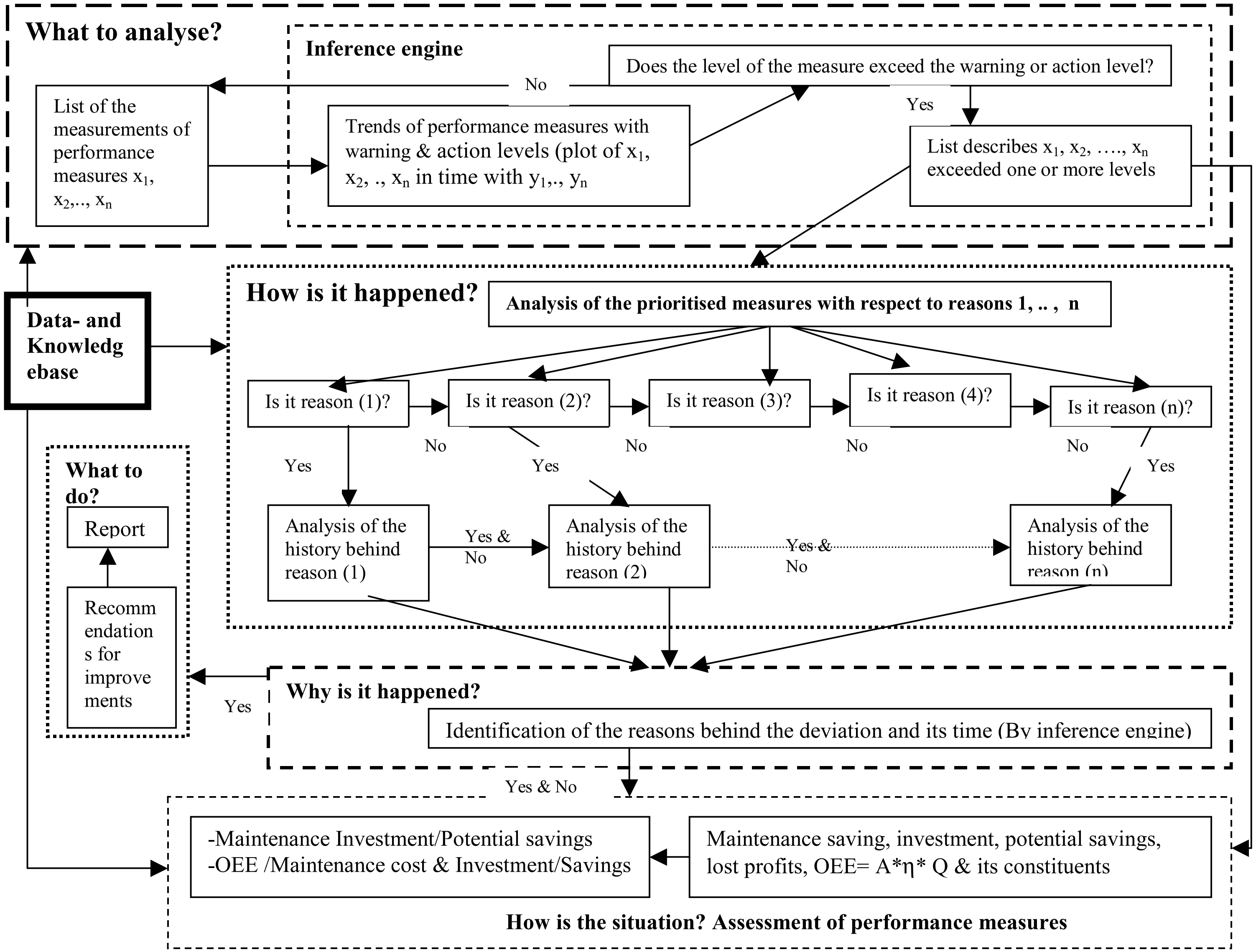


Fig. 1. A model for diagnosis of maintenance performance measure

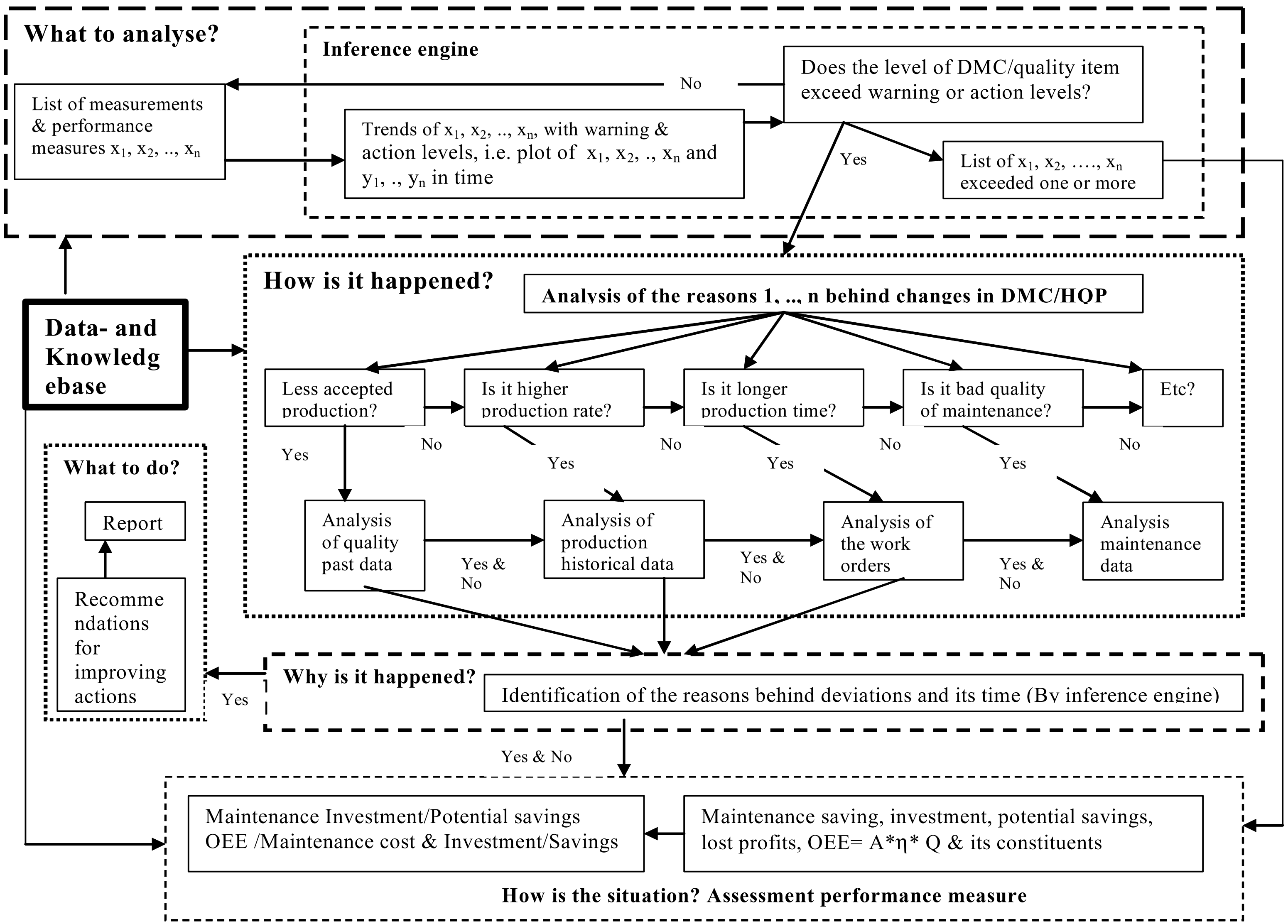


Fig. 2. Application of the model for diagnosis of maintenance performance measure when using maintenance direct cost divided by quality production

A SOCIO-TECHNICAL PERSPECTIVE ON INTEGRATED OPERATIONS FOR HIGH-RISK AND COMPLEX INDUSTRIAL ASSETS: EXPERIENCE FROM NORTH SEA OIL & GAS INDUSTRY

Jayantha P. Liyanage

Center for Maintenance and Asset Management, University of Stavanger, N.4036, Stavanger, Norway.

Abstract: During 2003 it became very obvious that oil and gas (O&G) exploration and production activities on North Sea was at a crossroad. Norwegian O&G industry during 2003-2004 made a strategic decision to step into a major re-engineering process of offshore activities. Subsequently the concept of 'Integrated eOperations' was born and is currently under a rapid implementation path. So far, the developments have mainly been high-tech driven, but it is widely acknowledged today that human and organizational aspects are the greatest challenges and hence some prudent and long-term measures are necessary to address them strategically. It implies that a more holistic socio-technical approaches are necessary to develop and implement high-performing systems given the level of complexity and the nature of change. This paper brings a brief overview of current developments on the Norwegian continental shelf, and outlines some of the critical aspects from a socio-technical perspective that are central for further advancement towards full-scale offshore-onshore integration by 2010.

Key Words: Smart assets, Asset management, Integrated eOperations, Socio-technical systems, eMaintenance, Asset performance, High-tech applications, Networking, Collaboration, Human and Organizational factors.

1 INTRODUCTION

This is an era when the global oil & gas (O&G) exploration and production activities are constantly challenged by various socio-political and techno-economical sources. Even though the supply-demand gap continuously extends owing to increased energy utilization trends, the harsh history of volatile economy particularly due to drastic reduction in oil price has substantially challenged major producers to draw more prudent operational strategies to enhance profitability and growth. Apart from security of economy, social and environmental uprise have induced various other challenges compelling major producers to re-calculate their risk and to manage uncertainties under complex operating conditions [1].

The current situation on North Sea activities is far more challenging and has already indicated the immediate need for a reengineering process. Various issues have been taken up for discussion over the last few years, ranging from taxation policies to new methods to improve recovery factor to implement feasible and effective solutions to overcome a range of unavoidable circumstances. The worries have been continuous since the uncertainties about the future have become more visible with direct impact on the risk exposure and the value creation. Among the major concerns there have been at least two issues in particular which has indicated that it is the time to 'get back to the drawing board' to engineer a completely new scenario for the entire Norwegian Shelf. These two issues are;

- Declining production
- Rising lifting costs

They have triggered a re-engineering process across the entire Norwegian Oil & Gas (O&G) industry, and subsequently O&G production and exploration activities on the Norwegian Continental Shelf (NCS) are at a major turning point today.

2 MAJOR DEVELOPMENTS ON 'NCS'

As the thorough analysis of the situation convinced the executive boards and the senior policy makers that 'value potential' on NCS is extremely high but the risk exposure of conventional practice is substantial, they began to look into how emerging complexities, uncertainties, and risks can be reasonably managed through smart solutions. Subsequently, the *Norwegian continental shelf (NCS)* made a strategic decision to step into a new integrated asset management scenario termed *Integrated eOperations*. The long-term plan has been on more strategic capitalization of digital technologies and more prudent use of operational networks and collaborative partnerships to establish a completely new operational setting (see Figure 1). For more information see [2], [3], [4], [5], [6], [7].

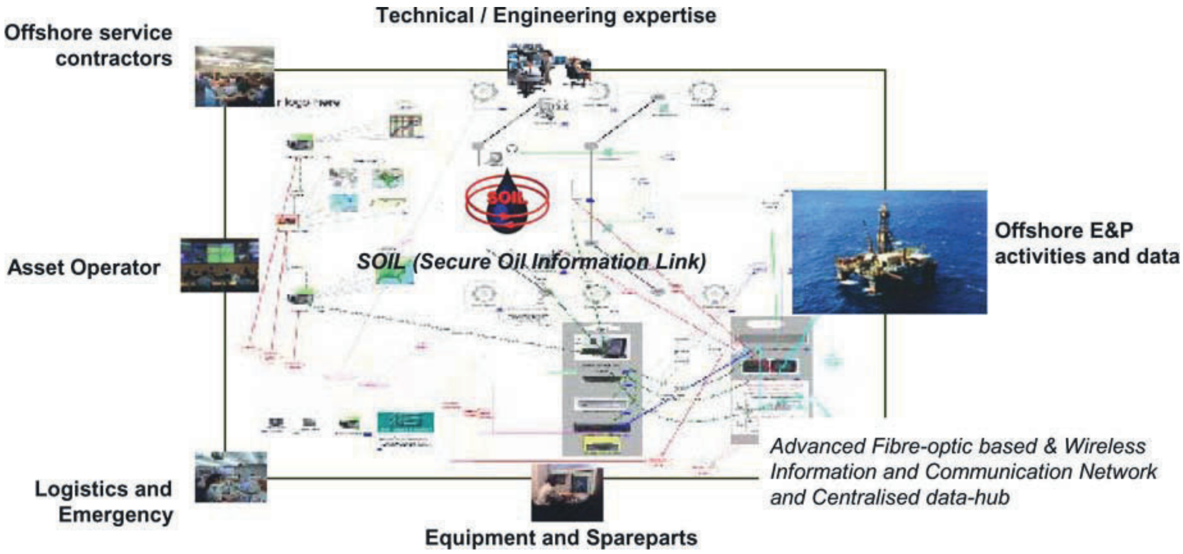


Figure 1. New asset management environment in North Sea.

This new practice has already begun to redefine and re-engineer the conventional setting by establishing online and real-time connectivity and interactivity between offshore production plants and onshore support environments. The traditional support system has been extended to establish more strategic collaborative partnerships with the supply chain and expert services using large-scale IT platforms. In principal, this looks into systematic offshore-onshore integration through application of advanced ICT technologies, and thus building high reliable information and knowledge based networks to remotely manage complex, high-risk, and capital intensive offshore assets, regardless of the geographical location. Ideally the new environment is expected to re-engineer the traditional infrastructure based on new forms of partnerships and cooperation, shared responsibilities and roles, integration of competencies, 24/7 connectivity, real-time accessibility to knowledge and information, and active communication networks. Current developments mainly target building up dedicated and reliable network of discipline experts and organizations, who are actively involved in dynamic sharing of information and knowledge and online real-time communication for fast yet effective decisions and actions based on 24/7 online real-time data exchange solutions (see Figure 2).

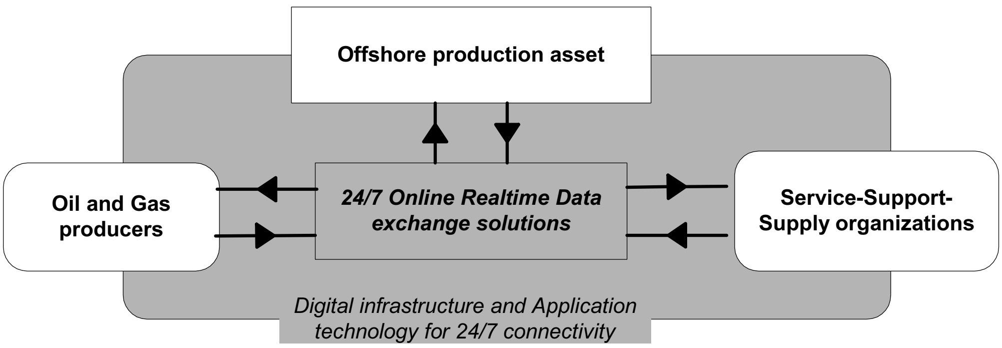


Figure 2. Interactive and connected environment for offshore asset management through smart technical solutions.

The new set up is expected to bring substantial commercial benefits in terms of;

- more commercially attractive offshore oil production,
- improving recovery or production regularity,
- more positive health, safety and environmental results, and
- improving economical performance, particularly by reducing operating costs

3 THE EMERGING SOCIO-TECHNICAL ENVIRONMENT

The major implication of new developments on NCS is that it will introduce a completely new socio-technical setting within the Norwegian O&G industry. Following the much earlier work done by Trist & Bamforth [8] and the Tavistock Institute, Socio-technical systems model in general views organizations as transformative agencies (see for instance [9]). The change dynamics of such agencies rests by far on the relational dynamics between personal, technical, and environmental (internal as well as external) sub systems. It portrays that humans and technologies in conjunction with a particular work system setting, that comprises certain organizational structures, routines, procedures, and work processes, involves in a continuous exchange process with the external environment on which the survival and success depends. Under the current circumstances on the NCS, and certainly in respect of the value creation potential, the conventional socio-technical boundaries are largely challenged and subsequently are subjected to expansion. It implies that the dynamic change largely occurs in such a way that more formal system settings, which have been largely centralised within specific organizational boundaries, are expanded on a networked and a collaborative basis to create an extended operational enterprise based setting. Connections between various actors of the new system is purely high-tech enabled and data & knowledge dependent. The transition towards new socio-technical setting is not going to be a smooth process, particularly owing to;

- complexity
- scale of change
- nature of exchange between different actors

The enhancing complexity is a direct a consequence of advanced application technologies and expanding organizational boundaries that substantially complement developments within techno-organizational space. Such complexity has direct internal organizational and external socio-political ramifications. At the very organizational level, the tendency to induce significant changes particularly within human and work-process related aspects has already become an obvious issue. At the more macro scale, this has also become an issue of national interest owing to safety and security concerns that emerge as a consequence of advanced open data management solutions and web-based collaboration.

The term ‘change’ is at the heart of almost every aspect related to the new development scenario. Changes can occur at various scales ranging from necessary adjustments towards establishing new competence profiles to major ones such as common data exchange platforms between actors based on XML, WITSML languages, O&G ontologies, semantic web, etc. It implies that initiatives and programs for smart change management is extremely critical to avoid negative consequences and serious losses, which may be too costly for the entire Norwegian O&G industry.

The conventional practice of exchange between different actors is largely being challenged, and the timely need for completely new forms of collaborative models has constantly been underlined by major stakeholders. It implies that the nature of exchange between different actors currently undergoes necessary major revisions to optimise collaboration between offshore assets, asset owners, and third party stakeholders that includes engineering, supply, knowledge-based, industry. New collaborations are mainly technology-enabled and web-based, and has brought an array of important aspects such as open data exchange platforms, performance incentives, risk and gain sharing models, organizational trust, etc. for ongoing debates and discussions. It has become very obvious so far that apart from very conventional commercially driven collaboration, the influence from non-commercial collaboration is too high to ignore. Such non-conventional yet critical collaborative models have to be established particularly with trade unions, trade organizations, authorities, and universities and higher educational institutes.

The new socio-technical environment will not be established within a year or two, but the current target is to realize full-scale operations by 2010. The transition will occur through 2 principal stages termed ‘G-1’ and ‘G-2’;

- G-1: Systematic integration between different disciplines. More cross-disciplinary coordination of planning and work execution. ICT applications remain more within defined organizational limits with localised technologies such as VisiWear.
- G-2: Systematic integration between different organizations. More cross-organizational coordination of planning and work execution. ICT applications are more expanded beyond organizational boundaries with more standardised web-based solutions.

As much as there are dedicated ambitions at whatever level, major actors need clear visions, prudent strategies, and smart operational plans to realize their objectives. At the current pace of development, three specific conditions are notable and certainly the entire industry needs more cautiously designed plans for further advancement with necessary checks and balances to ensure that all critical interfaces are well considered and analysed prior to implementation of solutions. Those conditions are;

- ambitions are set at very high levels
- time-scale to realize objectives is substantially reduced, and
- programs are accelerated within organizations to be champions of the new development

While development of technology and its successful implementation are challenges themselves, a major need is also to resolve those critical issues at organizational level that can induce major resistance and act as change impediments. A critical general concern both at socio-political and industrial levels is that ill-defined interfaces and increasing complexities of systems and data solutions can lead to unforeseen consequences, greater vulnerability, and greater risk. This mainly demands further attention to a great extent on the critical interfaces between human, organizational, technical, and work process related aspects in the design of high-performing social-technical systems for Integrated eOperations to realize full-scale benefits of reengineering efforts.

4 SOME CRITICAL SUCCESS FACTORS IN TERMS OF SAFETY AND SECURITY

Reaping the major benefits of Integrated eOperations relies heavily on every possible measure to avoid conventional socio-technical systems design practices that has largely contributed historically towards dys-functional or mal-functional systems development and modification efforts. Hendrick [10] for instance identifies 3 of such flawed design practices, namely;

- Technology-centred design: engineering aspects and technologies take the precedence with improper integration of operators, maintainers, and organizational settings for design, development, and implementation tasks. Particularly owing to cost and time constraints, more humanised, user-friendly, and organizationally compatible approaches are largely limited leaving ‘off the shelf’ or ‘cook book’ type engineering solutions demanding a major changes within human and organizational components at the usage phase.
- ‘Left-over’ approach: this is often a result of high-tech focus where owners, operators, and maintainers do have to take over the challenge of finding solutions for those issues that are left-over by the designers. It implies that design flaws, which largely arise by treating the personnel and organizational components as impersonal elements of the dynamic system, need to be taken care by the users and maintainers at later stage through ‘quick-fixes’ and ‘ad-hoc’ solutions.
- Mal-integration of an organization’s specific socio-technical characteristics: this situation arises when designers fail to fully comprehend the dynamic interactivity of socio-technical components within an organizational setting. The specific relationships, customs, end-user needs and comfort within the organizational setting are left largely ignored resulting in only sub-optimal system development and implementation.

A cautious and a well thought through process is absolutely necessary for NCS today, not purely from financial standpoint but also due to critical safety and security reasons. In relation to current stage of development of leading O&G producers, some of those issues that are considered critical for front-end engineering tasks are highlighted in Table 1.

Any re-engineering effort that is known to introduce major changes within a work system need to adapt much broader criteria to adequately consider key characteristics of the resulting socio-technical system. Sound priory knowledge and understanding of the dynamics of the system is critical for an effective and efficient system design and implementation task. At the end, the principal need is to ensure a more humanised and an organizationally oriented practice. Human factor professionals recommend three such specific criteria to support system design, i.e. joint design, humanised task approach, and organizations’ socio-technical characteristics integration (see [3]).

- Joint design: following a design practice where technological and personnel sub-systems are brought together at very early stages on a concurrent basis rather than a serial basis.
- Humanised task approach: more early and cautious considerations of function and task allocation process for a more humane work system.
- Organizations’ socio-technical characteristics integration: early assess the key socio-technical characteristics of an organization and properly integrate them into the new work system.

It has become more and more obvious at this moment that a comprehensive overview of the new system is necessary to identify sub-disciplines to properly integrate human and organizational aspects to the technological revolution on NCS. Proper set of procedures, routines, and guidelines are necessary to perform analysis, design, implementation, and evaluation tasks professionally to retain the desired level of integrity and consistency across the entire industry.

5 CONCLUSION

Since the high-level policy makers and trade organizations became aware that commercial risks of O&G production and exploration activities on NCS would increase substantially during the next few years, offshore activities on the continental shelf has been undergoing a major re-engineering process. The Norwegian O&G industry today is on a fast track along Integrated eOperations that is widely acknowledged as the best option to capture a substantial portio of remaining value of major reserves on NCS and to prolong the commercial life cycle by 30+ more years beyond 2030-2040. Major initiatives have already been taken by major sources of influence, and NOK billions of more investments will be made during the next few years for further advancement. While high-tech solutions have been a major driver for ongoing developments, it is widely acknowledged that human and organizational aspects perhaps pose as the greatest challenge for further advancement. It implies that there is a need for a proper detailed understanding and knowledge on the critical socio-technical characteristics of the emerging environment to ensure that the industry will not be exposed to unnecessary risks due to poor designing, decision making, and implementation planning. There is also a timely need to adapt more humane and organizationally-oriented approaches as opposed to those for instance as ‘technology-centred’ and ‘left-over’ that are historically known to be dys-functional for new systems development and implementation efforts. This obviously is going to be a major challenge during the forthcoming phases of the development activities.

Organisational change	
Rules and regulations	*Applicable rules and regulations*
Participation process	*Involvement of operators and maintainers in the change process*
Motivation	*Principal motivational factors for employees and partner organizations*
Trust in own role	*Building trust in relation to different roles in a new operating environment*
Communication	*Establish far better communication methods using advanced technologies*
Management	*Necessary support, infrastructure, and incentives to manage change*
New organization	*New socio-technical system based on interaction and collaboration*
Individual and production aspects	*Individual roles and responsibilities in respect of productivity and results*
Function analysis and location	*Role and task specification between offshore and onshore organizations*

Work environment	
Rules and regulations	*Applicable rules and regulations*
Psycho-social	*Psycho-social conditions of various work situations in offshore and onshore*
Effects on operations	*Effects of work environment changes on the day to day operations*
Effects on Production & Safety	*Effects of new work environment on production and safety performance*
Teams versus working alone	*Social relationships*
Changed status	*Nature and scale of change*
Individual adaptation	*Ability of individuals and teams to absorb the change and to perform right tasks*
Leadership role	*Coaching and mentoring capabilities, and quality of leadership*

Decision making, activity planning, and work coordination	
Rules and regulations	*Applicable rules and regulations*
Normal and emergency	*Management of normal vs. emergency operations*
Team responsibility	*Responsibility delegation patterns and mechanisms*
Risk perception	*Abilities to perceive risk exposure during decision making and work execution*
Trust	*Trust building measures for among employees and between partner organizations*
Process proximity	*Fast and quick responses for calls from offshore and onshore centres*
Competence (short and long term)	*Means to build new competencies*

Human – Technology interaction	
Rules and regulations	*Applicable rules and regulations*
Control room technology (Design)	*User-enabled design of offshore control rooms and onshore support centres*
Collaborative technology	*Implementation of user-enabled collaborative technologies*
Information presenting and processing	*Humanised information management capability*

Table 1. Some examples of important socio-technical issues for front-end-engineering tasks of Integrated eOperations on NCS.

6 REFERENCES

[1] Liyanage JP, (2003) Operations and maintenance performance in oil and gas production assets: Theoretical architecture and Capital value theory in perspective, *PhD thesis*, Norwegian University of Science and Technology.

[2] Liyanage JP, (2005) Managing Integrity of Offshore Assets through Digital Capability: Reducing Risk and Adding Value through Integrated OMS in North Sea, In *Proceedings of the 4th International Conference on Quality and Reliability*, Beijing, China. pp 193-199.

[3] Liyanage JP, (2005) Reducing risk and Adding Value through Smart Integrated Assets: Managing Complex Industrial Assets in Complex Environments, In *Proceedings of the 18th International Congress and Exposition on Condition Monitoring and Diagnostic Engineering Management,* Cranfield, UK. pp 123-130

[4] Liyanage JP, (2004) Smart integrated OMS: The 3rd leap to manage the integrity of high risk, capital intensive, and technologically complex offshore assets on NCS, In *Proceedings of the 4th Asia Pacific Conference on Systems Integrity and Maintenance*, New Delhi, India, pp. 409-414

[5] Liyanage JP, (2004) Digital future of operations and maintenance on the Norwegian Continental Shelf, In *Proceedings of the 17th International Congress and Exposition on Condition Monitoring and Diagnostic Engineering Management,* University of Cambridge, Cambridge, UK, Aug.-2004., pp 476-486

[6] Liyanage, JP, Herbert M, & Harestad J, (2006) Smart integrated e-operations for high-risk and technologically complex assets: Operational networks and collaborative partnerships in the digital environment, In Wang, Y.C., et.al., (ed.), *Supply chain management: Issues in the new era of collaboration and competition*, Idea Group, USA. (*forthcoming*)

[7] Liyanage JP, Langeland T, (2006) Smart assets through Integrated eOperations: Managing high-risk assets in complex environments through digital capabilities, In Mehdi Khosrow-Pour (ed.), *Encyclopedia of Information Science and Technology*, Idea Group, USA. (*forthcoming*)

[8] Trist EL & Bamforth KW, (1951) Some social and psychological consequences of the longwall method of coal-getting, *Human Relations*, 4, 3-38.

[9] DeGreene KB, (1973) *Sociotechnical systems*, Englewood Cliffs, NJ: Prentice Hall.

[10] Hendrick HW, (1995) Humanizing re-engineering for true organizational effectiveness: A macroergonomic approach, In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*, Santa Monica, CA, Human Factors and Ergonomics Society.

STRATEGIES FOR SUSTAINABILITY RISK MANAGEMENT THROUGH ACTIVE *'BAP (BUSINESS-ASSET-PROCESS) PERFORMANCE INTEGRATION'*

Jayantha P. Liyanage

Center for Maintenance and Asset Management (SDV), University of Stavanger, N-4036, Stavanger, Norway

Abstract: This paper sheds some light on the strategies for successful *BAP performance integration* based on knowledge and experience from oil & gas (O&G) industry. It defines and outlines an empirical framework to manage *sustainability risks* through systematic integration of performance across core business, asset portfolio, and core-processes such as operations & maintenance (O&M). It underlines two specific issues of timely significance. Firstly, O&M has a specific role-play towards a sustainable future when the global industrial activity shows an explosive growth, and secondly within sustainable business frameworks O&M process has an important value-adding role.

Key Words: Sustainability risk, Performance, Performance integration, Asset management, Operations and Maintenance, Performance quality, Performance measurement.

1 INTRODUCTION

Sustainability performance today has become an investable concept that effectively and efficiently balances profits and fundamentals for competitive advantage in complex and dynamic environments. It is a forethought business approach dedicated to create long-term value through better embracement of opportunities, reduction of sustainability costs, and management of hidden risks, by paying heavy emphasis on economical accountabilities towards those who bear financial risks, well-being of societies and societal agents, and proper care towards the eco-system. Today it has become an important means to ensure long-term security of investments and to manage enterprise risk exposure in the current socio-economic and political environment (for more information see [1]).

Certainly, decisions and activities at the industrial production/manufacturing asset level and technical-process level have significant impacts on the *sustainability performance*. Even though the impact of *sustainability performance* on the commercial success of the business has become more clear today, there are no proper guidelines, procedures, or frameworks available for systematic integration of performance of high risk & complex industrial production or manufacturing assets and of critical technical processes such as operations & maintenance (O&M), with sustainable business performance criteria. It implies that active *BAP performance integration* process has become a critical and an inevitable requirement for competitive advantage and commercial success in today's industry, and that effective strategies are necessary to steer the integration task.

2 SUSTAINABILITY CONCEPT

The World Commission Report on Environment and Development entitled *Our Common Future* issued in 1987 by Brutland Commission (see [2]) is an important landmark for the growth of the new idea of 'responsible' business. It has given birth to years of debates and discussions, and in principal has brought sustainability issues into limelight as an important concept for the survival and well-being offering a unique vision to achieve economic growth through responsible business practices.

Ever since, various activities and programs around the world by major sources of socio-economical influence, for instance UN, Dow Jones [3] etc. have contributed to bring two issues of interest to the attention of business leaders, i.e. *corporate social responsibility* and *sustainable business*. In fact the origin of both concepts remains more or less the same, and it can be widely seen today that these concepts are used in quite similar contexts without much of a difference in their meanings. In general, corporate social responsibility as a business principle represents a marriage of profits and fundamental socio-oriented principles; principles those emerged consequential to rapid globalisation requiring a far-reaching balance between business, government, and society at large. It covers range of a company's interaction with society ranging from conditions of employment, industry and labour standards, to social development and human rights, etc. It is believed to play an effective role as a strategy that fits with current industrial circumstances for gaining competitive advantage in the eyes of the societies particularly when business activities expand into more hostile regions. In the present context of business applications the distinction between corporate social responsibility and sustainable business is that, while the former mainly rests on the societal impact of corporate performance, the later seek for a blend of economic prosperity, environmental quality, and social equity, and captures a much broader scope and presents a composite picture of a legitimate value-creating business.

In essence, for O&G business, *sustainable business activity* implies; *adapting business strategies and activities that meets the needs of its wider stakeholders and the organization in question focusing on protecting, sustaining, and enhancing human and natural resources for future so that business activities are not only profitable but also importantly remains legitimate in the eyes of those bulk of stakeholders.*

In general this more broadly encompasses:

- social progress that equally recognize, respect, and cater to the needs of societal agents and organizations
- efficient and effective means to protect the eco system,
- cautious and prudent use of natural resources for the benefit of everyone, and
- contribute and sustain high and stable levels of economic growth of those who bear stakes

3 SUSTAINABILITY RISKS AND GAINS

The majority of O&G producers today seems to recognize that sustainable business policies and principles is the pathway to better deal with business uncertainties and risks, and the best option for competitive advantage (see for instance [4]). In essence, sustainability risk encapsulates;

- economical exposure
- corporate reputation
- partner goodwill

Sustainability performance has gained an added momentum ever since shareholders and other venture capitalists have taken serious interests to include similar criteria in the calculation of investment risks. This is highly commended for instance by *Dow Jones Sustainability Index* (DJSI) and policies for *Socially Responsible Investments* (SRI). There appear to be a rapid growth of interest in the investment community to follow the guidelines and policies of DJSI and SRI. Notably furthermore, non-governmental organizations and indigenous communities have begun to take legal steps against illegitimate business exploration and production activities in various parts of the world with serious economical consequences for O&G producers.

The pay back for implementation of prudent strategies to manage sustainability risk cannot simply be projected against purely quartile economical targets. It implies that business frameworks for sustainability risk management is about developing an expression about an organization's commitment towards its stakeholders and defining a clear mandate to achieve competitive advantage aiming a relatively longer-term success. Some of the contributing factors to competitive advantage include [1];

- reduced hidden costs
- created investment options
- attraction of investments
- reputational dividends
- gains in customers
- captured talent
- positive influence on product and service innovation
- creative leaderships, enhanced strategic relationships, etc.

However, it is known that the existing economical theories are not capable of sensibly expressing a major portion of the benefits in pure financial terms.

4 THE PERFORMANCE INTEGRATION CHALLENGE

While the executive leadership set guidelines and policies, and specify objectives abstractly that need to be accomplished at business level (see [5]), at the asset management level it is expected that operations are streamlined to substantiate those specifications and thus business improvements and developments. This implies that there should be an effective and an efficient correspondence between what the asset portfolio does (i.e. performance) and business results, and that there should be criteria to define the link between the two. O&G production assets have a specific role in the sustainable O&G business. In fact this role has to be complimented by various decisions and actions taken across assets at various levels and core processes. Asset processes should adapt performance management practices that allow them to realize their designated mission that directly relates to the asset performance criteria. In offshore O&G exploration and production environment 4 of such core asset processes exist;

- Drilling and well interventions
- Development and modifications
- Operations and maintenance
- Logistics and other support services

Notably, every core process has a pre-defined set of roles and accountabilities on the asset. For instance, *Drilling and Well intervention* largely dedicates to *recovery efficiency*, while *O&M* is more associated with assuring *technical condition* of equipment and systems and *safety integrity* of the asset. It implies that each process has to establish an appropriate framework in relation to its designated asset-based role to optimise its own performance.

Today there is an absence of standard criteria to define operational and technical specifications that explicitly explain what specific issues must be considered at the asset and asset-process levels to comply with sustainable performance. It implies that there is an inherent gap in the performance integrity management process. This can mainly be attributable to inherent complexities of underlying issues and inconsistencies of technical language in use to systematically work towards technical specifications at tactical and operational levels.

5 STRATEGIES FOR SUCCESSFUL '*B-A-P PERFORMANCE INTEGRATION* '

The concept of *BAP performance integration*, as the name implies, underlines a strategic need to streamline industrial activities in production/manufacturing/process plants or facilities and its core-processes with specific sustainable performance frameworks of the business. It is about an attempt, within the business and the manufacturing/production/process asset environment in question, to understand and to express the complex causal phenomenon of performance in a sensible way and to ensure that decisions are taken and activities are conducted on a risk-informed basis, and that those decisions and activities are corporate value-driven.

A successful strategy for *BAP performance integration* process comprises a set of specific tasks in a specific order, i.e. (see Figure 1);

1. *first order task*: Define and develop sustainable performance framework of the business.
2. *second order task*: Identify strategic operational specifications for production/manufacturing/process assets and define asset performance acceptance criteria.
3. *third order task*: Identify technical performance specifications for asset processes and define quality-tracking criteria.
4. *fourth order task*: Define and develop criteria for periodical business impact assessment.

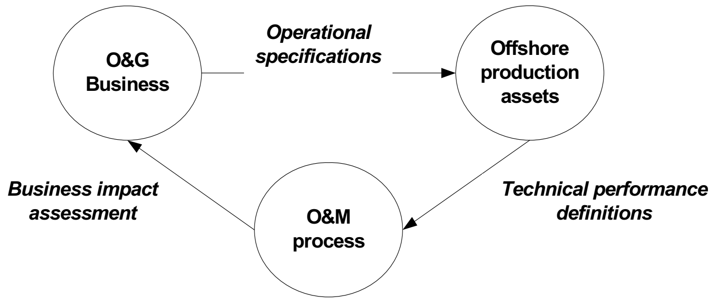


Figure 1. An illustration of *BAP performance integration* task (see [1], [6])

6 FROM FUNDAMENTALS TO DECISIONS AND ACTIONS

The popularising demand for more socially-responsible performance by the O&G business sector involves reconciling their legitimate search for profits with respect to rights and demands of stakeholders. This requires a re-orientation of O&G exploration and development paradigm based on rigorous and responsible assessment of trade-offs between temporary economic gains and the longer term pay back to stakeholders (Tomei, 1998). Such long-term business impact planning and subsequent sustainable business performance framework has to incorporate various social and environmental aspects. Figure 2 illustrates the guiding principals for developing such a framework in O&G business context.

As aforementioned a central challenge lies in the cascading of abstract business frameworks to operational specifications of the asset. It underlines the need of sensible criteria to define operational specifications that explicitly explain what specific issues must be considered at the asset level to comply with sustainable performance. If the asset owners comply with sustainable performance, then the ownership of an operating license of an offshore facility has to bring three-fold important issues into consideration, i.e.;

- *economical interests* (relates to direct financial interests of both the organization and those who bear financial risk, *e.g. actual production*)
- *liabilities* (directly address statutory and regulatory requirements imposed either by local or cross-border authorities, *e.g. safety integrity level*)

- *obligations* (encompass implicit moral and ethical aspects that are largely associated with exposure to vulnerabilities and harmful losses particularly during operational and decommissioning phases of high-risk assets, *e.g. waste handling*)

In that context, Table 1 and Figure 3 illustrate a generic criterion to establish relationships between sustainable business and asset performance.

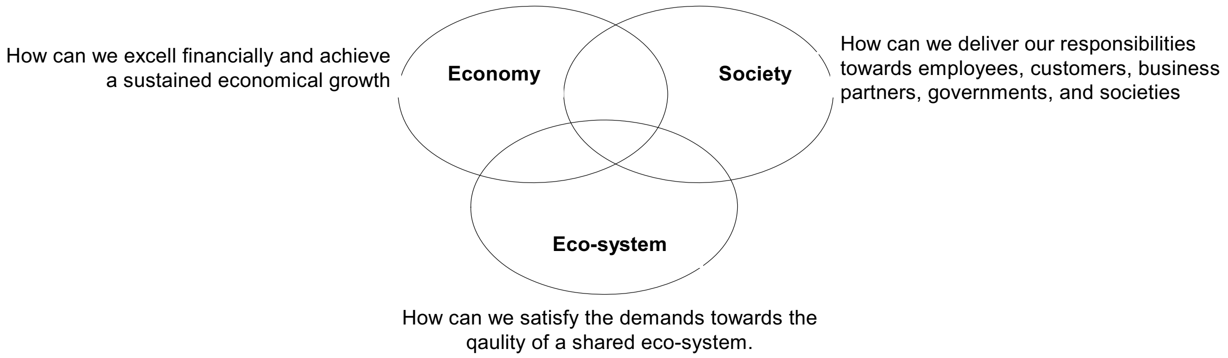


Figure 2. Guiding principals for sustainable business performance frameworks.

Table 1. Defining operational specifications.

	Sustainable Business		
Asset performance specification	**Economy**	**Society**	**Environment**
Economical interests	*e.g. actual production*	-	-
Liabilities	*e.g. insurance and compensations*	*e.g. working conditions*	*e.g. emissions, spills, discharges*
Obligations	*e.g. penalties*	*e.g. reduced health risk exposure and Quality assurance*	*e.g. waste handling*

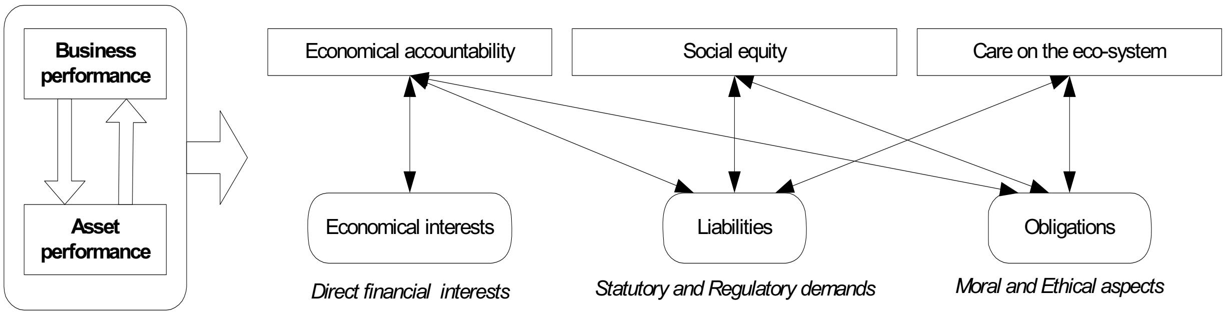


Fig. 3. Cascading from *Sustainable business performance* to *Sustainable asset performance*

The other challenge lies in the task that asset operators/owners have to identify technical performance specifications to shape-guide performance of asset processes. This requires a clear definition of designated function of the process within the asset performance framework, i.e. what a specific process under consideration is supposed to deliver to the asset. For instance, O&M is the process involved in technically conditioning assets. O&M need to explore the landscape for decisions and activities and develop the best strategies to optimise its delivery performance to retain a specific technical condition that contributes to overall commercial vitality of assets.

The establishment of the O&M performance logic in relation to asset performance criteria is essentially a systematic top-down process. It commences by reviewing abstract performance objectives and systematically specifying and defining technical details through the preceding levels. Figure 4 illustrates the core steps in the definition of the logic that convert a set of top-down *principal questions* to a working model that explains bottom-up *causal relations*.

The constituent components are defined and the underlying performance structure is built-up through systematic expansion of the working logic. It implies that the constituent components, which are the building blocks of the performance structure, gradually emerge as the principal questions are directly addressed and systematically resolved. It calls for an attempt to seek for clear technical specifications highly relevant for a given setting.

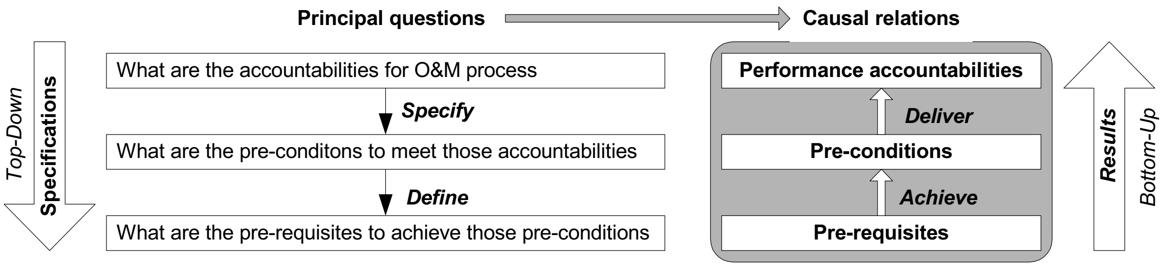


Fig. 4. Working logic to establish O&M performance framework (see [1], [6]).

Obviously, the actual detailed content of those *prerequisites* and formal organizational routines to manage those *perquisites* are fairly complex and hence remains beyond one's comprehension in its full scale. However, an attempt towards performance optimisation requires that such technical issues and adequately specified in relation to important circumstances within a given setting. A proper analysis and specification of technical issues will eventually result in establishing some form of a logic exploring interesting details related to;

- technical condition of the plant/facility
- safety integrity level of the plant/facility
- human factors
- organizational issues
- technologies and technology applications
- operational routines and work-processes

Adapting such a strategy largely helps performance optimization efforts as it helps avoiding specific performance biases. For instance, this particularly relates to abrupt slashing maintenance budgets for short-term cost savings, at the expense safety hazards and environmental damages. Such a framework brings positive effects to the O&M performance management efforts by providing a consistent basis for;

- key performance indicators and a balanced performance measurement system
- periodic performance audits or health checks
- quality assurance system for work processes, decisions and activities
- institution of internal standards and best practice documents for constant reference
- supportive infrastructure and a work setting

The central message here is that in relation to emerging concept termed *sustainable O&G business* and its implications at the production asset, and O&M process level, there is not only pure economical but also importantly legal, moral, and ethical bases for optimizing performance. It implies that the business case for O&M can be better expressed not only in terms of the cost of O&M activities, but also in savings in capital requirements, production regularity, occupational health, safety, and eco perspectives. This provides a stronger basis to express O&M as a *value-added process* rather than a pure *cost center* challenging the conventional thinking and practice. This has been discussed further by the authors elsewhere (see for instance, [7]-[11]).

7 CONCLUSION

This paper briefly discussed important issues related to the emerging concept termed *sustainable business* and strategies for *BAP performance integration*. The central argument brought into spotlight is that the business impact assessment of O&M performance, within emerging sustainable frameworks, should be performed not only in cost terms, but also looking into capital requirements, production regularity, occupational health, safety, and environmental results. It is stressed here that it is such a wider criteria that should be used as a basis model the critical business role of O&M process, particularly when complexities and uncertainties of business environment is too high to ignore and exposes the business to a wider set of risks in more dynamic socio-political settings.

8 REFERENCES

[1] Liyanage JP, (2003) Operations and maintenance performance in oil and gas production assets: Theoretical architecture and Capital value theory in perspective, PhD thesis, Norwegian University of Science and Technology, 2003.

[2] UN Department of Economic and Social Affairs, (2001) Sustainable Development, United Nations (UN).

[3] Dow Jones Indexes, (2001) Sustainability: Dow Jones Sustainability Group index, http://www.djindexes.com/jsp/index.jsp (A WWW Document).

[4] Armstrong K, (1994), Promoting sustainable development: Strategies for the exploration and production industry, Society of Petroleum Engineers, Article number SPE 27113, (www.spe.org).

[5] Wolff R, Zaring, O., et.al., (2000), Indicators for sustainable development, Society of Petroleum Engineers, Article number SPE 61320, (www.spe.org).

[6] Liyanage JP, (2005) Operations and Maintenance Performance Optimisation within Sustainable Business Frameworks: The Case for Complex and High Risk Assets, Proceedings of The 4th International Conference on Quality and Reliability (ICQR-2005), Beijing, China. pp 201-207.

[7] Liyanage JP, (2005) Sustainable O&M Performance Framework: Emerging Basis to Review and Re-define O&M Practice in High Risk and Complex Industrial Assets, Proceedings of the 18th International Congress and Exposition on Condition Monitoring and Diagnostic Engineering Management (COMADEM-2005), Cranfield, UK, pp 343-348.

[8] Liyanage JP (2004) Quality of operations and maintenance performance in sustainable production environments, 5th International Conference on Quality, Reliability, Maintenance (QRM-2004), University of Oxford, England, pp 175-178.

[9] Liyanage JP (2004) Sustainability business case for maintenance in offshore oil and gas production assets: Experience from North Sea, Proceedings of the 17th European Maintenance Congress: Euromaintenance 2004, Barcelona, Spain, May, 2004, pp 309-314.

[10] Liyanage JP, (2004) Operations and maintenance of complex assets in sustainable environments: The emerging basis for systematic management of quality, risk, profitability, and growth, The 4th Asia Pacific Conference on Systems Integrity and Maintenance (ACSIM-2004), New Delhi, India, pp 415-421.

[11] Liyanage, JP, Kumar U, (2003) Towards a value based view on operations and maintenance performance management, Journal of Quality in Maintenance Engineering, 9(4), pp 333-350.

THE NEED FOR AGILITY IN ASSET MANAGEMENT

Harris H [a] and Carapiet S [b]

[a] *Hawke Research Institute, School of Management, University of SA, North Terrace, Adelaide, SA 5000, Australia.*

[b] *School of Management, University of South Australia, North Terrace, Adelaide, SA 5000, Australia.*

Abstract: Increasingly turbulent environments faced by enterprises today have reduced the time available to organisations to prepare and respond to change. Organisational agility, characterized by a rapid and proactive response to opportunities and threats, is a source of competitive strength and adaptive capability for engineering enterprises. Asset management can be an effective source of competitive advantage if it occurs within the broader framework of sustained organisational agility. This paper demonstrates how an understanding of the concept of agility, and obtaining a simple measure of the levels required in an organisation, can help managers to incorporate agility into their asset management programs. The Agility Measure described in the paper draws on experience with a large multi-national project, SYMPHONY, in which many small and medium-sized enterprises participated. A key feature of the Agility Measure is its applicability in medium sized enterprises which tend to have limited tools available to managers.

Key Words: Asset management, Agility, Competitive advantage, SMEs, Adaptive enterprises

1 INTRODUCTION

Contemporary businesses, service organisations and government enterprises face an increasingly turbulent operating environment as a result of globalisation, advances in information technology, rising living standards, increased regulation and corporate governance demands, less cohesion in social values and the rise of post-modernism. Enterprises have responded to these challenges by vigorous pursuit of lower costs, increased clock speed, new organisation structures, increased participation in networks and clusters, and in some cases an involvement in virtual organisations. To be successful in this new environment firms, organisations and employees need capabilities and characteristics which may differ in nature from those needed in an earlier, more certain, age. Not only are additional technical skills needed but systems also may need to be more resilient, structures flatter and more flexible, 'soft' skills more prominent, reporting more open and the organisation as a whole more agile.

These changes impact on asset management activity. The paper shows how the changes in organisational theory and the business environment will impact on key features of engineering asset management. Operating in a turbulent environment has an impact on all aspects of a business, and there is no exception for management of assets – the 'continuous process covering the whole-of-life cycle of an asset from conceptual design through to construction [or] manufacture, operational use, maintenance, rehabilitation and/or disposal' [1]. Agile asset management, the ability to respond to rapid changes in the environment, offers an organisation competitive advantage for instance in product/process flexibility, customer response capability and ability to integrate new technology. We contend that the adoption in engineering asset management of the traits and attributes necessary for agile operation will bring benefits not only to the asset management process but also to the enterprise as a whole.

The paper begins by taking an historical look at the development of business systems and organisation structures, followed by an outline of the elements of chaos and complexity theory which underlie the new paradigm. It then outlines the key features of agile and adaptive organisations before discussing specific implications for engineering asset management.

2 THE EMERGENCE OF THE CONTEMPORARY ENTERPRISE

In the enlightenment of the eighteenth century Adam Smith provided a description of the way in which wealth is created [2] which underpinned a view of business and economics which was to hold sway almost unchallenged for over 200 years. Individual businesses, proprietors and entrepreneurs were to engage in brutal competition for profit, while an 'invisible hand' ensured that the overall economy operated efficiently. The more efficient producer would supplant the less efficient, and those goods which were the more needed in society would be produced to the exclusion of those less necessary.

Classical management theory has its roots in the growing demand for industrial products and the increase in the number of large enterprises which occurred as the industrial revolution gained pace. It was also firmly grounded in the scientific revolution and the Newtonian world view. *The Principles of Scientific Management* was published in 1911 by a mechanical engineer, F.W. Taylor, and the five functions of management – planning, organising, commanding, co-ordinating and controlling – were enunciated by the French engineer Henri Fayol in 1916 [3]. This led to a formal division of labour, even in the executive ranks, with distinctions between line and staff functions, where 'all the thinking is done by the managers and designers, leaving all the doing to the employees' [4]. Power within these bureaucratic organisations rested with fixed offices or roles, not with individual human beings. The role of the leader was to maintain equilibrium [5].

Aspects of this classical approach are evident still in many organisations today. The pursuit of efficiency has led to the appearance over recent decades of a series of techniques for enhancing the efficiency of organisations. Just-In-Time, avoiding waste through lean production, and the pursuit of best practice through benchmarking are just three examples.

It was not until the last half of the twentieth century that any serious challenge to the free market model emerged in the West. The dominance of the shareholder as the proper focus and recipient of business profit was first challenged [6] by the notion that others had a stake in the business. All stakeholders, including suppliers, customers and employees, were to be considered, not just the owners [7]. The 'Triple Bottom Line', which has achieved currency as a way to measure performance, requiring an assessment in each of the economic, social and environmental fields [8], further broadens the area of focus.

2.1 Importance of networks and virtual enterprises

One place where the overwhelming power of the simple, capitalist, view of the business as the unadulterated pursuit of profit in a dog-eat-dog competition has been progressively challenged is in the increasing use of networks, alliances and clusters in business. Maintenance of machinery may be outsourced to a specialist provider, former employees may form a cluster of self-employed contractors with premises on the former employer's site, and IT-based condition monitoring services may be provided by unseen people in another country. The formation of these new business groupings is widely encouraged yet there is less recognition of the significant shift in internal organisation [9] which is needed if networks are to deliver the benefits for which they were praised and taken up. The significance lies not only in the changed view of the purpose of business – co-operation with those who have been (and often still are) competitors is at the heart of the network model – but also in the management skills needed for success in such a business [10].

2.2 The nature of competition

Traditionally, competitive advantage has come from attributes such as location, operating cost and technology. Michael Porter identifies the threat of entry by new competitors, the intensity of rivalry among existing competitors, pressure from substitute products, the bargaining power of buyers, and the bargaining power of suppliers as the five forces that drive competition within an industry [11]. Another view focuses on the strengths and weaknesses of the business and the link between a firm's internal characteristics and its performance. Known as the resource-based model [12], it attributes sustainable competitive advantage to the firm's assets (or core competencies), with particular emphasis on those assets which are valuable, rare, not easily imitable and incapable of substitution [13].

On this view, not only are the physical resources – the rail networks, blast furnaces and large machines – a valuable source of competitive advantage, but so are human resources – as captured by the catch-phrase "our people are our greatest resource" – and also the operating procedures, corporate culture, reporting procedures and planning system. In engineering asset management this means that competitive advantage can come not only from the physical assets and the people who maintain them, but also from the systems used to manage the assets and their maintenance.

Taken together these changes represent such a large change that for those who accept them it has become a new paradigm, a new way of looking at the world [14]. In the following sections we show how this paradigm has emerged from the theories of chaos and complexity, and then show how agility is an important characteristic for enterprises which aim to succeed in this new world. This will set the stage for our discussion of the implications for engineering asset management.

3 A CHANGED VIEW OF THE WORLD: FROM NEWTON TO CHAOS

The Newtonian, scientific paradigm which underpins the traditional organisation structure viewed the world as a well-behaved machine. This has been challenged by the complexity paradigm where there are limits on the ability to predict future events [15-17]. Whilst there have been many best-selling books that have played a major role in enhancing the understanding and popularity of the theories of chaos and complexity [17-22] the relevance to management has seldom been explicit beyond a commitment to the replacement of the older, Newtonian model [23]. Three important elements are discussed below – the limits on prediction, the possibility of self-organisation and organisations as complex adaptive systems.

While the initial impetus for the research in complexity came from the natural sciences including chemistry, physics, biology and mathematics, numerous books aimed at management practitioners [17, 24-31] and the general public [20, 22] have raised awareness for the potential application of complexity theory. Complexity theory has been applied to management [5, 32, 33], self-organisation in entrepreneurial networks [34], self-managing teams [35], strategic processes [36], leadership [37, 38], innovation and knowledge management [39-41], development [42], and project management [43]. However there have been very few empirical studies [33, 41], and 'relatively little work on developing a theory of complex social systems', which is what organisations are [33].

3.1 Limits on prediction

The striking discovery that unpredictable behaviour is generated in simple deterministic systems with only a few elements is the foundation of what has come to be called chaos theory. When meteorologist Edward Lorenz discovered that 'prediction of the sufficiently distant future is impossible by any method, unless the present conditions are known exactly' [44], he termed this the "butterfly effect" - the notion that a butterfly stirring the air today in Peking can transform storm systems next month in New York. This places fundamental limits on prediction, even when there are few elements in the system and the relationship between them is precisely known. Tiny disturbances can produce exponentially divergent behaviour, which is not the case in Newton's universe.

3.2 Self-organisation

The second significant aspect of chaos theory is that 'astonishingly simple rules, or constraints, suffice to ensure that unexpected and profound dynamical order emerges spontaneously' [45]. The generation of new forms from inner guidelines, rather than by imposition from outside, is characteristic of self-organisation in open and living systems [46]. It is 'a process in which pattern at the global level of a system emerges solely from numerous interactions among the lower-level components of the system' [47]. The idea that order in the universe is natural and spontaneous [48] is dramatic and illuminating. Complexity theory takes the concept one step further, contending that organisation not only arises spontaneously but also is adaptive [49].

Communication is the key driver in the self-organisation process as it is the primary mechanism for interaction amongst the individuals (agents) in organisations (complex adaptive systems) [50-55], which leads (after Comfort [54]) to the characterisation of self-organisation as a collective process of communication, choice, and mutual adjustment in behaviour based on a shared goal among members of a given system without external order.

3.3 Organisations as complex adaptive systems

The dynamic ability of complex adaptive systems to evolve and learn over time can provide organisations with increased capability to handle adaptive challenges. Complexity should be seen not as 'a methodology or set of tools' but as a 'conceptual framework, *a way of thinking* and *a way of seeing the world*' [33] (emphasis in original). Viewing organisations as complex adaptive systems [37, 56-61] offers a powerful new lens to study and understand the dynamic emergence of order, structure or patterns from the interactions of the elements inside a system, whether that system be one in the natural world, or a business organisation [15, 57, 62]. Iterative learning within these systems can modify the rules for interaction as the environment changes and this allows the system to evolve and improve its performance in changing environments [63-65].

Self-organisation occurs in open systems that are far-from-equilibrium [66], that is in systems that are characterized by constant change and failure to reach equilibrium as distinct from ones following the Newtonian paradigm and moving towards equilibrium, thus making a manager a facilitator of self-organisation rather than a leader charged with the role of maintaining equilibrium.

Any system shows 'properties which are properties of the whole, rather than the properties of the component parts' [67], and a complex system is not one with many parts but one where the interaction between the parts is complex, while complex adaptive systems have the ability to adapt to the environment in a systematic way [49]. Human systems including organisations are adaptive non-linear feedback systems that exist under conditions far from equilibrium [4, 26, 57].

4 AGILE AND ADAPTIVE ORGANISATIONS

An organisation that can self-organize, a complex adaptive system as described in the previous section, has the internal capacity for spontaneously emerging structures, depending on what is required [17]. Organisations that tap into the property of self-organizing or self-renewing systems have been termed "adaptive organisations" where the task determines the organisational form [68]. An adaptive organisation has also been defined as being able to sense and respond to changes in its environment and learn from these responses to change its behaviour and evolve.

Another term used in the literature to describe organisations with these characteristics is 'agile'. Agile enterprises are described as dynamic and adaptive [69], able to operate profitably in a competitive environment of continuous and unpredictable change [70], and capable of gaining 'competitive advantage by intelligently, rapidly and proactively seizing opportunities and reacting to threats' [71]. The turbulence of the business environment reduces the response-time available to firms to cope with change and thus creates the need for agility. The key drivers of this need for agility are to be found in a company's business environment— such as change in the marketplace, technology, and competition [72, 73].

Agility is related to self-organisation in terms of survival in a turbulent environment. Some characteristics of agile companies bear similarity to self-organising firms. For example, the agile manufacturing reference model developed from research on agility in UK-based SMEs by the Agile Manufacturing Research Group [74] has four interdependent elements of agility – strategy, processes, linkages, and people. Agile people will be found where the organisation structure is adaptable, continuous learning occurs, people are multi-skilled and flexible, and decision making is rapid, based on robust, integrated systems. Closely associated with adaptable structures in this reference model is the concept of 'performing partnerships' in which agile organisations gain new and enlarged capabilities through cooperation, often linking together to form virtual enterprises [74]. Indeed the linking of firms to form virtual enterprises is a core component of the agile enterprise [75]. Thus, along with permeable boundaries, an acceptance of the unpredictability of outcomes, and effective communication between participants to maintain the relationship success in agile enterprises will require the trust, compatible management philosophies, informality and redundancy which are necessary for effective networks and alliances [10, 76].

5 IMPLICATIONS FOR ENGINEERING ASSET MANAGEMENT

Given that agility will be a benefit for an enterprise that faces a turbulent environment, what are the implications for those organisations which have substantial engineering assets and in particular for the management and maintenance of those assets? Thus far, we have shown that the benefits of agility will be compromised if only part of the organisation is agile while the rest continues in the pursuit of an excellence which is grounded in the certainty of the Newtonian model. Thus the engineering asset management system and those who run it cannot be divorced or isolated from the wider organisation. At the ICOMS conference in 2001 Harris [77] argued that the changes in corporate governance brought about by increased stakeholder interest in environmental and social performance would impact the reporting requirements and performance targets for maintenance management, and this paper extends that to include an even greater impact.

Any integrated approach to engineering asset management will included at least twelve elements [1], and one way to examine the impact that an acceptance of the new organisational paradigm will have is to consider how some of these elements will be affected. Three elements – data management, human resources, and condition monitoring – are considered below.

5.1 Data management

The process of asset management requires substantial information to be collected from many different parts of the organisation. Although organisations are generating more and more data, many managers lack confidence that they have enough reliable, consistent, correct and timely data for decision making. Organisations may have far more data than they use; yet not have the data they really need.

The 'soft factors' such as organisation structure, culture, and communication [78] have found to have a significant impact on data quality for engineering asset management which is in line with the ideas of the complexity paradigm. Communication between participants is the key driver of the self-organisation process and effective communication requires a shared language and values. This is one example of how the implications of complexity apply to engineering asset management.

Chaos theory offers companies that need to handle adaptive challenges a view of the world as dynamic where change is the norm not the exception, and where prediction is impossible. There are implications for organisations operating in a dynamic and unpredictable world not to follow the 'command and control' techniques of the past, as they would be likely to inhibit the ability to cope effectively with the changes. A much more viable management philosophy, given the existence of chaos, is to devise ways to embrace change and cope with change by having the capability to respond to unpredictable situations built into the organisation.

5.2 Human Resources

Enterprises which are organized on hierarchical principles, and that will include many asset management departments and maintenance units, require managers who are comfortable and effective in situations where command and control are key management functions. Complex adaptive systems deliver agility and competitive advantage but they require managers with different skills and attributes if they are to deliver these benefits. Technology and context-specific skills and knowledge remain important but the personal traits associated with success now include tolerance of ambiguity, inquisitiveness, and perhaps even love [10]. This is a particular challenge for those who have learnt their trade or gained their experience under an earlier, more certain regime, where organizations placed great store in "one right way".

In addition, as more organizations come to recognize that they are open systems the range of parameters by which performance is judged will broaden. Asset managers will be asked to report on environmental and social performance as well as availability and cost [77], further adding to the change for the manager brought up in the command and control, scientific management era.

5.3 Condition monitoring

Many consider condition monitoring to be 'the most cost-effective way' of maintaining assets as it 'enables deterioration in condition to be detected and remaining life to be predicted' based on regular inspection and measurement [79]. It relies on the predictability of future events. Conceptually, chaos theory cautions that no real life prediction can be relied on because the initial conditions are never fully known. This may be of little practical consequence for those working with the machine, in the same way that relativity and quantum theory have no discernible impact on the sizing of shafts or the measurement of rotation speed. Furthermore, the statistical elements which are found in many condition monitoring techniques reduce any tendency to consider the predictions as absolutely correct. For the researcher and theorist there may be more subtle influences which cannot be ignored.

These three examples are interlinked. Condition monitoring relies on data collection, and a human preference for certainty can encourage both recipients and providers of information to hide or gloss over the inherent unpredictability in it. Viewing an organisation as a complex adaptive system that is an open and non-linear system will lead to changes to the existing closed and linear system of asset management (and accounting). Making actual changes will not be easy, but unless they are made the benefits which agility and self-organisation can bring in response to the turbulent business environment will not be captured. An engineering asset manager has to accept the inability of maintenance systems to provide perfect prediction and work on developing the capabilities of agility, innovation and experimentation in his people and systems to operate in a dynamic world.

6 MEASUREMENT OF AGILITY

This section describes a measure of agility which managers can use to determine the need for agility in their organisations, the readiness of the organisation to become agile, and the extent of agility. The first prototypes of the Agility Measure were developed during a collaborative research project conducted under the Intelligent Manufacturing Systems (IMS) framework. Known as SYMPHONY, the project addressed the need of SMEs for management tools that assist in handling a turbulent environment and included 30 organisations from the European Union, Canada, Australia and Switzerland [80, 81]. Among the companies that participated in SYMPHONY were an Italian automotive parts and manufacturing company, a Swiss shipbuilder, an Australian provider of stormwater and wastewater filtration systems and a Swiss-based electricity distribution line support company.

The Agility Measure draws from the methodology for agility developed by Zhang and Sharifi [72, 73], and three reference models developed by practitioners [74, 82, 83]. The three modules of the Agility Measure provide assessments of the need for agility, the readiness for agility and agility level. The Agility Measure is a self assessment tool, with the first two modules principally designed for use by the manager or management team while the third module, which assesses the agility level of the enterprise, allows for data to be gathered from a wider range of participants, perhaps including those outside the organisation. In each module of the Agility Measure, managers see a list of evaluation questions in a number of categories, which they will need to rate on a scale of 1–5 (similar to a Likert scale). The individual characteristics which make up the assessment in each module have been chosen from the factors identified in the literature and reference models [72-74, 82-84] and are shown in Table 1. Although initially developed as a pencil and paper form, an online version is now available. Each module can be completed in less than 15 minutes.

Table 1

Components of the Agility Measure

Module	Categories included in the assessment	Number of items in each category
Need for agility	Marketplace	10
	Competition	4
	Customer requirements	4
	Technology	2
	Social factors	5
	Suppliers	2
	Internal complexity	6 total = 33
Readiness for agility	Inclination towards risk	3
	Previous experience with novel ideas	3
	Resources	4
	Political commitment	2
	Sophistication	4
	Intensity of rivalry between firms	3 total = 19
Agility level	Sensing, perceiving and anticipating changes	2
	Strategic vision	2
	Delivery speed and timeliness	2
	Time to market	2
	Cost effectiveness	4
	Operations efficiency and effectiveness	3
	People flexibility	2
	Knowledgeable, competent empowered people	2
	Product volume flexibility	3
	Rate of new product introduction	2
	Organisational flexibility	2
	Product model/configuration flexibility	3
	Business network	3
	Sufficient and appropriate technological ability	3 total = 35

The assessment of the need for agility module assists managers in deciding whether there is actually a need for their company to be agile. Essentially the degree of turbulence in the business environment determines the need for agility. The questions in this module assess the paradigm in which the manager is operating. Agility is needed in enterprises that operate in an unstable and unpredictable environment, as in the complexity paradigm.

The assessment of readiness for agility module assists managers in answering the question: how ready is my company for agility? The questions in this module are about the experience of the enterprise with change as a response to the far-from-equilibrium conditions it has faced. The past experience of the enterprise and its earlier achievement in adaptation are reflected in this assessment of readiness for agility. This module relates to the concept of organisations as complex adaptive systems that learn and evolve, based on their experiences. As in the case of self-organisation, the enterprises that learn from their experiences tend to improve their performance over time.

The assessment of agility level module assists managers in assessing the current level of agility of their company. The flexibility of the enterprise in terms of its people, processes, and linkages with other enterprises is part of the agility level of an enterprise. A high level of agility corresponds to a strong adaptive capability, and the indicators captured in this module include the anticipation of change, strength of the company vision, and sharing of information openly.

The measures can be customised to individual organisational requirements in two ways. The measurement proformas each include a section where the degree of importance of each category for the company in terms of its performance success is determined. The respondents simply rate each category as being of low, medium, or high importance. The application of these weightings to the raw scores for each category ensures that the overall score for the module is tailored to reflect the importance of the various categories for the business. In addition enterprise-specific questions can be added, for instance on specific elements of cost effectiveness or to make reference to particular aspects of the environment.

7 CONCLUSION

Globalisation, demands for more stringent and transparent corporate governance, and advances in information technology mean that many businesses, government instrumentalities and service providers face a more turbulent operating environment. These changes have been accompanied by changes in the way in which many people look at the world. As the industrial revolution was accompanied by great changes in science and religion so the concepts of chaos and complexity are increasingly associated with the management and structure of contemporary enterprises. In particular some enterprises are seeking to emulate the capability for self-organisation which is has been identified in many systems in the natural world so that these human organisations might become more agile and better able to adapt to the rapidly changing environment. Those responsible for engineering asset management cannot stand aside from these changes, which require an acceptance of uncertainty, permeable boundaries, high levels of trust and effective communication to support relationships and learning. The Agility Measure described in the paper provides managers with a tool for assessing the need for agility in their organisation or unit, together with the organisation's readiness and agility level.

8 REFERENCES

[1] CIEAM. (2005) *CIEAM - Asset management innovators.* Brisbane: CRC for Integrated Engineering Asset Management.

[2] Smith A. (1970) *The Wealth of Nations.* Harmondsworth: Penguin Books.

[3] Fayol H. (1949) *General and Industrial Management*. London: Pitman.

[4] Morgan G. (1986) *Images of organization*. Beverly Hills: Sage Publications.

[5] Tetenbaum TJ. (1998) Shifting paradigms: from Newton to chaos. *Organizational Dynamics*, 26(4), 21-32.

[6] 6.Berle AA & Means GC. (1968) *The Modern Corporation and Private Property.* New York: Harcourt, Brace & World.

[7] Freeman RE. (1998) A stakeholder theory of the modern corporation. In LB Pincus Hartman (Ed) *Perspectives in Business Ethics*. pp.171-81. Chicago IL: Irwin/McGraw-Hill.

[8] Elkington J. (1998) Cannibals with forks: The Triple bottom line of 21st century business. Gabriola Island BC: New Society.

[9] Drucker PF. (1995) *Managing in a time of great change*. London: Butterworth-Heinemann.

[10] Harris H & Bititci U. (2005) The importance of moral values to the effectiveness of networking in Manufacturing and Maintenance. *New Engineer*, 8(1), 5-10.

[11] Porter ME. (1985) Competitive advantage: creating and sustaining superior performance. New York: The Free Press.

[12] Wernerfelt B. (1984) A Resource-based View of the Firm. *Strategic Management Journal*, 5, 171-180.

[13] Barney JB. (1991) Firm resources and sustained competitive advantage. *Journal of Management*, 17(1), 99-120.

[14] Kuhn TS. (1970) *The structure of scientific revolutions*. Chicago IL: University of Chicago Press.

[15] McMillan EM. (2004) *Complexity, organizations and change*. London: Routledge.

[16] Nicolis G & Prigogine I. (1989) *Exploring complexity: an introduction*. New York: W.H. Freeman.

[17] Wheatley MJ. (1994) Leadership and the new science: learning about organization from an orderly universe. San Francisco: Berrett-Koehler Publishers.

[18] Gell-Mann M. (1995) The quark and the jaguar: adventures in the simple and the complex. London: Abacus.

[19] Gleick J. (1988) *Chaos: making a new science*. London: Cardinal.

[20] Merry U. (1995) Coping with uncertainty: insights from the new sciences of chaos, self-organization, and complexity. Westport, CT: Praeger.

[21] Ulrich H & Probst GJB, eds. Self-organization and management of social systems: insights, promises, doubts, and questions. 1984, Springer-Verlag: Berlin.

[22] Waldrop MM. (1994) Complexity: the emerging science at the edge of order and chaos. London: Penguin.

[23] Mitleton-Kelly E. (2004) An Integrated Methodology to Facilitate the Emergence of New Ways of Organising, In (Ed) *Proceedings of 3rd European Conference on Research Methodology of Business and Management Studies (ECRM),* Reading University

[24] Ralls Jr. JG & Webb KA. (1999) The nature of chaos in business: using complexity to foster successful alliances and acquisitions. Houston, TX: Gulf Professional Publishing.

[25] Haeckel SH. (1999) Adaptive enterprise: creating and leading sense-and-respond organizations. Boston, MA: Harvard Business School Press.

[26] Goldstein J. (1994) The unshackled organization: facing the challenge of unpredictability through spontaneous reorganization. Portland, OR: Productivity Press.

[27] Watson TJ. (1994) In search of management: culture, chaos and control in managerial work. London: Routledge.

[28] Peters TJ. (1989) Thriving on chaos: handbook for a management revolution. London: Macmillan.

[29] Axelrod R & Cohen MD. (1999) Harnessing complexity: organizational implications of a scientific frontier. New York: Free Press.

[30] Wood R. (2000) Managing complexity: how businesses can adapt and prosper in the connected economy. London: The Economist Books.

[31] Nonaka I & Takeuchi H. (1995) The knowledge-creating company: how Japanese companies create the dynamics of innovation. New York: Oxford University Press.

[32] Ulrich H & Probst GJB, eds. Self-organization and management of social systems: insights,promises, doubts, and questions. 1984, Springer-Verlag: Berlin.

[33] Mitleton-Kelly E. (2003) Ten principles of complexity and enabling infrastructures. In E Mitleton-Kelly (Ed) *Complex systems and evolutionary perspectives of organizations: the application of complexity theory to organizations*. pp.23-50. Amsterdam: Elsevier.

[34] Biggiero L. (2001) Self-organizing processes in building entrepreneurial networks: A theoretical and empirical investigation. *Human Systems Management*, 20(3), 209-222.

[35] Molleman E. (2000) Modalities of self-managing teams: the "must", "may", "can", and "will" of local decision making. *International Journal of Operations & Production Management*, 20(8), 889-910.

[36] Zimmerman B. (1993) The inherent drive towards chaos. In P Lorange, et al. (Ed) *Implementing Strategic Processes: Change, Learning and Co-operation*. pp.373-393. Oxford: Blackwell.

[37] Regine B & Lewin R. (2000) Leading at the edge: how leaders influence complex systems. *Emergence*, 2(2), 5-23.

[38] Marion R & Uhl-Bien M. (2003) Complexity theory and Al-Qaeda: examining complex leadership. *Emergence*, 5(1), 54-76.

[39] Andriani P & Romano A. (2001) Introduction: distributed systems of knowledge and complexity theory. *International Journal of Innovation Management*, 5(2), 129-134.

[40] McElroy MW. (2000) Integrating complexity theory, knowledge management and organizational learning. *Journal of Knowledge Management*, 4(3), 195-203.

[41] Kitchell S & Wilfrid LU. (1995) Corporate culture, environmental adaptation, and innovation adoption: a qualitative/quantitative approach. *Journal of the Academy of Marketing Science*, 23(3), 195-205.

[42] Rihani S. (2005) Complexity theory: a new framework for development is in the offing. *Progress in Development Studies*, 5(1), 54-61.

[43] Jaafari A. (2003) Project management in the age of complexity and change. *Project Management Journal*, 34(4), 47-57.

[44] Lorenz EN. (1963) Deterministic non-periodic flow. *Journal of the Atmospheric Sciences*, 20(2), 130-141.

[45] Kauffman SA. (1995) At home in the universe: the search for laws of self-organization and complexity. New York: Oxford University Press.

[46] Loye D & Eisler R. (1987) Chaos and transformation: implications of non-equilibrium theory for social science and society. *Behavioral Science*, 32(1), 53-65.

[47] Camazine S, et al. (2001) *Self-organization in biological systems*. Princeton, NJ: Princeton University Press.

[48] Kauffman SA. (1995) *At home in the universe*. New York: Oxford University Press.

[49] Frederick WC. (1998) Creatures, corporations, communities, chaos, complexity: a naturological view of the corporate social role. *Business & Society*, 37(4), 358-389.

[50] Simon HA. (1976) Communication. In (Ed) Administrative behavior: a study of decision-making processes in administrative organization. pp.154-171. New York: Free Press.

[51] Jantsch E. (1980) The self-organizing universe: scientific and human implications of the emerging paradigm of evolution. Oxford: Pergamon Press.

[52] Hammond SC & Sanders ML. (2002) Dialogue as social self-organization: an introduction. *Emergence*, 4(4), 7-24.

[53] Olson EE & Eoyang GH. (2001) Facilitating organization change: lessons from complexity science. San Francisco, CA: Jossey-Bass.

[54] Comfort LK. (1994) Self-organization in complex systems. *Journal of Public Administration Research & Theory*, 4(3), 393-410.

[55] Katz D & Kahn RL. (1966) *The social psychology of organizations*. New York: John Wiley and Sons.

[56] Boisot M & Child J. (1999) Organizations as adaptive systems in complex environments: the case of China. *Organization Science: A Journal of the Institute of Management Sciences*, 10(3), 237-252.

[57] Stacey RD. (1996) *Complexity and creativity in organizations*. San Francisco, CA: Berrett-Koehler Publishers.

[58] Morel B & Ramanujam R. (1999) Through the looking glass of complexity: the dynamics of organizations as adaptive and evolving systems. *Organization Science: A Journal of the Institute of Management Sciences*, 10(3), 278-293.

[59] Harkema S. (2003) A complex adaptive perspective on learning within innovation projects. *The Learning Organization*, 10(6), 340-346.

[60] Fuller T & Moran P. (2001) Small enterprises as complex adaptive systems: a methodological question? *Enterpreneurship & Regional Development*, 13, 47-63.

[61] Pascale RT. (1999) Surfing the edge of chaos. *Sloan Management Review*, 40(3), 83-94.

[62] Mitleton-Kelly E. (2000) Complexity: partial support for BPR? In P Henderson (Ed) *Systems Engineering for Business Process Change*. pp.24-37. London: Springer-Verlag.

[63] Axelrod NN. (1999) Embracing technology: the application of complexity theory to business. *Strategy & Leadership*, 27(6), 56-59.

[64] Tan J, Wen HJ & Awad N. (2005) Health care and service delivery systems as complex adaptive systems. *Communications of the ACM*, 48(5), 36-44.

[65] Andersen P. (1999) Complexity theory and organization science. *Organization Science*, 10(3), 216-232.

[66] Nicolis G. (1989) Physics of far-from-equilibrium systems and self-organization. In P Davies (Ed) *The New Physics*. pp.316-347. Cambridge: Cambridge University Press.

[67] Checkland P. (1999) Soft Systems Methodology: a 30 year retrospective. Chichester: John Wiley.

[68] Dumaine B. (1991) *The Bureaucracy Busters.* p 37-50.

[69] Ashkenas RN. (1995) The boundaryless organization: breaking the chains of organizational structure. San Francisco: Jossey-Bass Publishers.

[70] Goldman SL, Nagel RN & Priess K. (1995) Agile Competitors and Virtual Organizations: Strategies for Enriching the Customer. New York: Van Nostrand Reinhold.

[71] Bessant J, et al. (2001) Developing manufacturing agility in SMEs. *International Journal of Technology Management*, 22, 28-54.

[72] Zhang Z & Sharifi H. (2000) A methodology for achieving agility in manufacturing organisations. *International Journal of Operations & Production Management*, 20(4), 496-512.

[73] Sharifi H & Zhang Z. (2001) Agile manufacturing in practice - Application of a methodology. *International Journal of Operations & Production Management*, 21(5/6), 772-794.

[74] Meredith S & Francis D. (2000) Journey towards agility: the agile wheel explored. *The TQM Magazine*, 12(2), 137-143.

[75] Carrie AS. (1999) Integrated clusters: the future basis of competition. *International Journal of Agile Management Systems*, 1(1), 45-50.

[76] Elmuti D & Kathawala Y. (2001) An overview of strategic alliances. *Management Decision*, 39(3), 205-217.

[77] Harris H. (2001) Excellence in Maintenance: Have the Goalposts Moved? In P Pascoe and J Kennedy (Ed) *Proceedings of International Conference of Maintenance Societies,* Melbourne ppPaper 012:1-8

[78] Koronios A, Lin S & Gao J. (2005) A Data Quality Model for Asset Management in Engineering Organisations. Cambridge MA.

[79] Kinsky R. (1994) *Engineering Management*. Melbourne: Nelson.

[80] Kemp JLC, et al. (2004) An integrated set of methods for strategic management of knowledge-based, adaptive SMEs, In M Taisch (Ed) *Proceedings of International IMS Forum 2004: Global Challenges in Manufacturing,* Cernobbio IT pp1236-1243 Politecnico di Milano

[81] Merino E, et al. (2004) Competences management for dynamic collaborative value-creating networks [IMS Symphony project], In M Taisch (Ed) *Proceedings of International IMS Forum 2004: Global challenges in manufacturing,* Cernobio Italy pp1291-1296 Politecnico di Milano

[82] Dove R, Hartman S & Benson S. (1996) An Agile Enterprise Reference Model with a Case Study of Remmele Engineering. Agility Forum.

[83] Canter J, *An agility-based OODA model for the e-commerce/e-business enterprise*. marchFIRST. http://www.belisarius.com/modern_business_strategy/canter/canter.htm. (2000).

[84] Harris H, Carapiet S & Provis C. (2004) Adaptive and agile organizations: do they actually exist? *Philosophy of Management: formerly Reason in Practice*, 4(1), 3-12.

9 ACKNOWLEDGEMENTS

The authors have benefited from discussion of the concept of agile organisations and the application of chaos and complexity theory to human organisations at a number of conferences including the 7th Stimulating Manufacturing Excellence in SMEs conference in Glasgow and at various events associated with the SYMPHONY project. The Australian Government through the Innovation Access program assisted the attendance of Australian participants at SYMPHONY meetings. Saadia Carapiet's doctoral studies in self-organisation in SMEs has been supported in part by scholarships from Smartlink: National Institute for Manufacturing Management and the Cooperative Research Centre (CRC) for Integrated Engineering Asset Management (CIEAM).

MAINTENANCE RELATED IEC DEPENDABILITY STANDARDS

Per Anders Akersten[a]

[a] *Centre for Dependability & Maintenance, and Division of Operation & Maintenance Engineering*
Luleå University of Technology, SE-97187 Luleå, Sweden

Abstract: Within the International Electrotechnical Commission, IEC, there exists a committee, named IEC Technical Committee 56 "Dependability", TC56 for short. Its history can be traced back to 1965, when the Reliability and Maintainability Committee was formed. In 1990, after contacts with ISO, it was decided that the IEC/TC56 should be responsible for standardization in the field of dependability, in any appropriate technological area, including those normally considered as outside the scope of IEC. The term "dependability" is a collective term, used to describe availability performance and its constituent factors reliability performance, maintainability performance, and maintenance support performance. Within the IEC/TC56 a number of maintenance related standards and guidelines have been produced, covering e.g. maintainability of equipment, reliability-centred maintenance, integrated logistic support, and also an application guide entitled "Maintenance and maintenance support". Also, Chapter 191 of the International Electrotechnical Vocabulary, freely available on the Internet, contains a wealth of maintenance-related terms. This paper is intended to give an overview of the standards available and some insight into the ongoing efforts within IEC/TC56 to produce new standards and to maintain existing ones. It should not be seen as an official promotion of the IEC Dependability Standards series. Instead, it presents some personal experiences and reflections on the use and usefulness of the standards series.

Key Words: standardization, dependability, management, maintenance, maintainability, life cycle cost, logistic support, ILS, RCM

1 INTRODUCTION

The International Electrotechnical Commission (IEC) Technical Committee 56 "Dependability", TC56 for short, performs standardization work within dependability. Its history can be traced back to 1965, when the IEC Reliability and Maintainability Committee was formed. Today, the scope of TC56 is formulated as follows [1]:

"To prepare international standards in the field of dependability, in all appropriate technological areas, including those not normally dealt with by IEC Technical Committees. ... The standards provide systematic methods and tools for the dependability assessment and management of equipment services and systems throughout their life cycle."

The term "dependability" [2, (IEV 191-02-03)] is a collective term, used to describe availability performance and its constituent factors reliability performance, maintainability performance, and maintenance support performance. The various dependability standards are related to product issues, process issues as well as management issues. They do not cover safety, although their applications often address both dependability and safety issues.

Of special interest within maintenance management are the published IEC-standards describing

- Life Cycle Costing (LCC)
- Maintainability
- Reliability Centred Maintenance (RCM)
- Integrated Logistic Support (ILS)
- Maintenance and Maintenance Support

These standards contain general introductions to concepts and provide guidance for applications within different areas. They do not describe detailed procedures and should not be compared to the technical standards, having narrower scopes, covering specific products or approaches.

Different standardization issues have been addressed at previous ICOMS Conferences, see e.g. [3], [4] and [5]. This paper is intended to give an overview of the dependability standards available and some insight into the ongoing efforts within IEC/TC56 to produce new standards and to maintain existing ones. It should not be seen as an official promotion of the IEC Dependability Standards series. Instead, it presents some personal experiences and reflections on the use and usefulness of the standards series.

2 THE STANDARDS PRODUCTION PROCESS

Within the work of ISO and IEC TC 56 is classified as a horizontal Technical Committee. This means that it should exclusively treat fundamental principles, concepts, terminology or technical characteristics that are relevant to a number of other horizontal or product Technical Committees. The other committee type, product Technical Committee, has a scope, covering a specific product or group of related products.

Today TC 56 has 33 member countries, 23 so-called P-members, and 10 so-called O-members. P-members are supposed to participate actively in the work, to vote on all questions formally submitted for voting within the technical committee or subcommittee, and to participate in meetings. O-members are allowed to follow the work as observers and have the right to submit comments and to attend meetings.

The TC 56 Dependability consists of four working groups:

- WG 1. Dependability terminology
- WG 2. Dependability techniques
- WG 3. Dependability management
- WG 4. System aspects of dependability

These groups constitute the four Maintenance Teams and they are responsible for the establishing and managing of Project Teams. Moreover, the Committee has liaisons with several other IEC and ISO committees, e.g. ISO/TC 108/SC 5 "Mechanical vibration and shock – Condition monitoring and diagnostics of machines" and ISO/TC 176 "Quality management and quality assurance".

The production of standard documents proceeds through a number of stages. The IEC and ISO standards production processes are very similar, see e.g. [4]. For the life of IEC standards the following stages can be identified:

Table 1. Stages in the life of an IEC standard

Stage		**Activity**	**Related document types**
Project stages	Preliminary	Introduction of new items, subject to voting among P-members.	PWI (Preliminary work item)
	Proposal	Proposition of a new work item, preparation of first working draft, nomination of project leader.	NP (New work item proposal)
	Preparatory	Setting up of project team, preparation of working draft	WD (Working drafts)
	Committee	The principal stage, containing preparation of committee drafts, circulation of drafts and consideration of comments from national bodies.	CD (Committee drafts)
	Enquiry	Circulation of enquiry draft, voting and submission of comments by national bodies, revision (if needed) of document.	CDV (Committee draft for vote)
	Approval	Distribution of final draft, voting by national bodies.	FDIS (Final draft international standard)
	Publication	Correction of errors (if any) in final draft, printing and distribution of standard.	IEC Standard
Maintenance		Examining need for revision of existing standards, establishing of revision work as new projects.	From CD to updated IEC standard.
Withdrawal		Withdrawal of existing standards.	

3 EXISTING MAINTENANCE-RELATED DEPENDABILITY STANDARDS

Appendix A provides a list of existing international standards, prepared by TC 56 Dependability and related to maintenance. The standards can be seen as organized in four different levels:

1. The top level documents (60300-1) deal with dependability management system and dependability programs.
2. The second level documents (60300-2) give guidance on various program elements and tasks.
3. The third level documents (60300-3) are so-called application guides.
4. The fourth level documents describe different tools, such as procedures and statistical techniques.

The standards related to maintenance management can be found among the third level documents, dealing with

- Life cycle costing (60300-3-3)
- Maintainability (60300-3-10)
- Reliability centred maintenance (60300-3-11)
- Integrated logistic support (60300-3-12)
- Maintenance and maintenance support (60300-3-14)

The series of standards relating to maintainability have been around for some time, and they do not fit perfectly into the four-level scheme. They may be seen as application guides, although they in several cases describe useful tools in more detail.

IEC 60300-3-3 "Life cycle costing" provides an introduction to the concept of life cycle costing, with special reference to costs associated with the product's dependability. The document describes the value of life cycle costing and outlines general approaches. It also covers common life cycle cost elements and guidance for performing a life cycle cost analysis.

IEC 60300-3-10 "Maintainability" is an application guide for maintainability, to be used in the implementation of a maintainability program for the life cycle phases *initiation*, *development* and *in-service* of a product. It gives guidance on how maintenance aspects are considered in design and modification work in order to achieve the appropriate maintainability.

IEC 60300-3-11 "Reliability centred maintenance" provides guidelines for the use of reliability centred maintenance (RCM) analysis techniques in the development of preventive maintenance programs for an equipment or a structure. The present version of the standard is based largely on the procedures in MSG-3. However, it is applicable not only to aircrafts but to a variety of items.

IEC 60300-3-12 "Integrated logistic support" describes the process of the management method ILS, by which the logistic support services required by a customer can be brought together in a structured way. These logistic support services include maintenance, manpower, training, spares, documentation, packaging, handling, storage, transportation, support resources and disposal. The standard provides guidance on minimum activities necessary for an effective ILS implementation.

IEC 60300-3-14 "Maintenance and maintenance support" describes a framework for maintenance and maintenance support and also minimal common practices. It further describes related management, processes and methodologies. The approach is more general than the one used in integrated logistic support (ILS) and is applicable also to cases where maintenance and maintenance support have to be adapted to specific conditions during the various life phases of a system.

It should be noted that several of these application guides are related to the previously issued standard IEC 60706-4 "Guide on maintainability of equipment - Part 4 - Section 8: Maintenance and maintenance support planning".

Another maintenance-related application guide is under preparation. According to the New work item proposal (NP) document, its tentative title is IEC 60300-3-16 "*Guideline for the specification of maintenance support services*" and its scope *"This IEC standard describes a framework for the specification of services related to the maintenance support of products, systems and equipment that are carried out during the operation and maintenance phase. The purpose of this standard is to outline, in a generic manner, the development of agreements for maintenance support services as well as guidelines for the management and monitoring of these agreements by both the company and the service provider."*

4 MAKING USE OF THE STANDARDS

Standards may be used in different ways. They are not laws, but they can be used as a basis for contracts, for evaluation of providers of service, for risk and insurance management, for decisions regarding contract compliance, and possibly also for related litigation and prosecution. The compliance to the IEC dependability standards is voluntary and the benefits from using them have to be identified. The benefits, technical as well as strategic, are very well understood by the experts participating in the standardization work. One main problem is to make decision-makers aware of the strategic and other potential benefits from the use of international standards.

The standards represent a massive amount of knowledge, compiled and condensed by a balanced group of international experts. The published standards also represent a certain degree of consensus (more than simple majority) of the group of experts, as well as of the voting members of the Technical committee, i.e. the participating countries' national committees. This can be utilized by letting the appropriate maintenance related standards constitute a core of a body of knowledge within maintenance management and engineering. This facilitates the dissemination of knowledge to companies, educational organizations and other organizations.

Among other benefits of the use of international standards, product oriented as well as management oriented, can be mentioned, e.g., the following.

- They are used world-wide
- They can be used by companies and organizations, regardless of size
- They provide a common language for description of processes

- They provide checklists of the main issues to be covered
- They provide a framework, facilitating performance measurement and improvement work
- They facilitate education and training within its subject area
- They reduce technical barriers to trade by harmonizing of national standards

Another way of making use of international standards is to participate in the standardization work. The homepage of Standards Australia [2], describes the value of participating with the use of the headings *Value to you*, *Value to your business* and *Value to the nation*. Those of us who have had the opportunity of participating immediately recognize the values *Personal satisfaction, Networking opportunity (national as well as international), Knowledge gain, Familiarity with standards, and Recognition for your organization.*

This section of the paper is closed by a quote from Valter Loll [7], convenor of TC56 WG2: *"If you want to stay updated you have to participate in the standardization work or receive information from someone who does."*

5 DISCUSSION

The use of standards in training and education for emergency management is discussed in [8], and a list of twelve requirements on a standard is provided. Although the application area discussed in the paper is quite narrow, most of the requirement types are generally valid. A few examples are as follows:

- *A standard should clearly define the terms it relies on for explanation.*
- *A standard should aim to be acceptable to as large a body of users as possible.*
- *A standard should seek to homogenise terminology, methods, procedures etc in order to reach consensus about common objectives.*
- *A standard should not be didactic, i.e. it should not seek to train.*
- *A standard should not inhibit the development of higher levels of professionalism.*
- *A balance must be struck between the need for permanence and the need for revisions.*
- *A standard should be freely available and distributed without cost to all who may benefit from it.*

Looking back into the history of the IEC dependability standards we may observe several occasions where some of these requirements were not fulfilled. Some individuals have strongly criticized the standardisation work and the standards published [9] and also comments from national bodies have had an impact on how the committee work is performed today. Most of the requirements listed above are now fulfilled to a large extent. One exception is the last requirement. The IEC offers free access to the vocabulary [2] but not to the individual standard documents.

For the effective use of international standards in the development of maintenance management and maintenance technology we need to stay informed. There exist several standardization bodies beside IEC and ISO; in Europe you will find, e.g., CENELEC and CEN, and in the United States you will find ANSI and IEEE. Moreover, most industrial nations have national standardization bodies. The different actors on the standardization arena do cooperate to a great extent, but the layman needs help to find out about existing standards within a particular subject. The establishment of a comprehensive and updated list of standards related to maintenance and product support should be added as a work item for the national maintenance societies as well as for future ICOMS.

6 REFERENCES

[1] IEC SMB/2936/R (2004). TC 56 Dependability: Strategic Policy Statement. Genève: IEC.

[2] IEV (International Electrotechnical Vocabulary). http://std.iec.ch/iec60050 (Accessed January 2006).

[3] Dunn, S. (2005). The maintenance dictionary: a plea for common maintenance terminology. In: Proceedings of ICOMS 2005, Hobart, Tasmania.

[4] Hitchcock, L. (2005). Evolving international standards for machine condition monitoring. In: Proceedings of ICOMS 2005, Hobart, Tasmania.

[5] Jamieson, D. (2005). Upcoming international standards on tribology-based machine condition monitoring. In: Proceedings of ICOMS 2005, Hobart, Tasmania.

[6] Standards Australia, http://www.standards.org.au (Accessed January 2006.)

[7] Loll, V. (2001). Internatinal standards for reliability. Quality and Reliability Engineering International, 7, 362.

[8] Alexander, D. (2003). Towards the development of standards in emergency management training and education. Disaster Prevention and Management, 12(2), 113-123.

[9] O'Connor, P. (1998). Standards in reliability and safety engineering. Reliability Engineering and System Safety, 60(2), 173-177.

6.1 Appendix A: published maintenance-related iec dependability standards

General guidelines

IEC 60300-3-1, Ed.2 (2003). *Dependability management - Part 1: Dependability management systems.*

IEC 60300-3-2, Ed.2 (2004). *Dependability management - Part 2: Guidelines for dependability management*

IEC 60300-3-1, Ed.2 (2003). *Dependability management - Part 3-1: Application guide - Analysis techniques for dependability - Guide on methodology*

Maintenance management, application guides

IEC 60300-3-3, Ed.2 (2005). *Dependability management - Part 3-3: Application guide - Life cycle costing*

IEC 60300-3-10, Ed.1 (2001). *Dependability management - Part 3-10: Application guide – Maintainability*

IEC 60300-3-11, Ed.1 (1999). *Dependability management -- Part 3-11: Application guide -- Reliability centred maintenance*

IEC 60300-3-12, Ed.1 (2001). *Dependability management - Part 3-12: Application guide - Integrated logistic support*

IEC 60300-3-14, Ed.1 (2004). *Dependability management - Part 3-14: Application guide - Maintenance and maintenance support*

Maintainability, a number of tools

IIEC 60706-1, Ed.1 (1982). *Guide on maintainability of equipment. Part 1 - Sections One, Two and Three. Introduction, requirements and maintainability programme*

IEC 60706-2, Ed.1 (1990). *Guide on maintainability of equipment. Part 2 - Section Five: Maintainability studies during the design phase*

IEC 60706-3, Ed.1 (1987). *Guide on maintainability of equipment. Part 3 - Sections Six and Seven. Verification and collection, analysis and presentation of data*

IEC 60706-4, Ed.1 (1992). *Guide on maintainability of equipment - Part 4 - Section 8: Maintenance and maintenance support planning*

IEC 60706-5, Ed.1 (1994). *Guide on maintainability of equipment - Part 5: Section 4: Diagnostic testing*

IEC 60706-6, Ed.1 (1994). *Guide on maintainability of equipment - Part 6: Section 9: Statistical methods in maintainability evaluation*

6.2 Appendix B: maintenance-related iec ongoing standardization work

NB: This list is compiled in January, 2006. Work items introduced later are not included.

IEC 60300-3-1, Ed.2 (2003). *Dependability management - Part 1: Dependability management systems.*

IEC 60300-3-11, Ed. 2. *Dependability management - Part 3-11: Application guide - Reliability centred maintenance*

IEC 60300-3-15, Ed. 1. *Dependability management - Part 3-15: Guidance to engineering of system dependability*

IEC 60300-3-16, Ed. 1. *Dependability management - Part 3-16: Application guide - Guideline for the specification of maintenance support services*

IEC 60706-2, Ed. 2. *Maintainability of equipment - Part 2: Maintainability requirements and studies during the design and development phase*

IEC 60706-3, Ed. 2. *Maintainability of equipment - Part 3: Verification and collection, analysis and presentation of data*

IEC 60706-5, Ed. 2. *Maintainability of equipment - Part 5: Testability and diagnostic testing*

IEC 62402, Ed. 1. *Obsolescence management - Application guide*

E-MAINTENANCE AND VULNERABILITY

Per Anders Akerstena

[a] *Centre for Dependability & Maintenance, and Division of Operation & Maintenance Engineering*
Luleå University of Technology, SE-97187 Luleå, Sweden

Abstract: Teleservice, utilizing modern information processing and communication tools, is by now a well-known way of supporting a product and we have seen the emergence of e-manufacturing and related types of e-service, in particular e-maintenance. In discussions of the present and future of e-manufacturing and e-maintenance systems are among other things risks and vulnerabilities related to utilization of e-services identified as an important area for further research. In this paper the concepts of e-maintenance and vulnerability are briefly described. The vulnerability of the e-maintenance system and its consequences for the managed assets are discussed to some extent. From different risk analysis approaches, e.g. related to computer and network security, several methodologies and tools for the management of vulnerability can be identified and an analysis methodology is proposed. The paper represents the view of a risk analyst and reliability engineer.

Key Words: e-maintenance, e-service, vulnerability, risk analysis, security

1 BACKGROUND

Teleservice, utilizing modern information processing and communication tools, is by now a well-known way of supporting a product, see e.g. [1]. It has been in use a number of years, e.g. in aviation, manufacturing, robotics and also in medical supervision and treatment. In recent years the concept of product support has been discussed in more detail, and the need for the combination of support to product as well as support to client has been clearly identified, see e.g. [2] and [3]. This development can be seen in the emergence of e-manufacturing and related types of e-service, in particular e-maintenance. There is a continuous development of the approach, as well as of related methodologies and tools – technical and organizational. In the book chapter [4] Koc et al discuss the present and future of e-manufacturing systems and e-maintenance. Among other things they point out the risks and vulnerabilities related to utilization of e-services as an important area for further research.

This paper discusses to some extent the vulnerability of the e-maintenance system and its consequences for the managed assets. After having performed a number of literature searches, using information portals such as Scopus, IEEE Explore and Scirus and reference databases such as Inspec and Compendex, I have identified very few research papers directly related to e-services and risk or vulnerability issues. Nevertheless, the problem area is important and has been recognized within neighbouring areas, such as telemedicine [5] and telematics [6].

The aim of the present paper is to briefly discuss the concept of vulnerability. From different risk analysis approaches, e.g. related to computer and network security, several methodologies and tools for the management of vulnerability can be identified. The paper represents the view of a risk analyst and reliability engineer.

2 THE CONCEPT OF E-MAINTENANCE

In the paper [7] Koc and Lee give the following definition of an intelligent e-maintenance system: “Intelligent Maintenance System (IMS) is an internet-based and web-enabled predictive maintenance technology which consists of intelligent machine degradation assessment, e-prognostics, and e-diagnostics to enable manufacturers and customers to have products and production machines with near-zero-breakdown conditions. Remote and real time assessment of machine’s performance information requires an integration of many different technologies including sensory devices, reasoning agents, wireless communication, virtual integration and interface platforms.”

The e-maintenance system consists of several different elements, e.g. sensors, analysing and reasoning agents, communication channels and man-machine as well as machine-machine interfaces. The system can be described by using a hierarchy of levels, from product or process level via system level to global, world-wide, level. Among its functions can be mentioned

- Real-time monitoring of machine behaviour and health
- Fault diagnostics
- Machine performance degradation assessment and prognostics

- Predictive maintenance
- Exchange of information (product/process level as well as system level) via appropriate information networks. This includes the sharing of information related to similar products or processes between different partners and geographical locations.
- Extraction of information from huge amount of data, data mining
- Provide business related decision support to the user, enabling the optimization of asset management

In [3] the concept of e-manufacture has been described and the relation between the e-manufacture systems and e-maintenance systems clarified. The e-maintenance system gives opportunities for the establishment of new types of relationships between the supplier of the product (or service) and the customer, enabling high quality service during the product's entire life.

3 THE CONCEPT OF VULNERABILITY

The broad term dependability covers several aspects of a system's ability of offering service to the user. The formal definition, taken from the International Electrotechnical Vocabulary [8] is "*The collective term used to describe the availability performance and its influencing factors: reliability performance, maintainability performance and maintenance support performance.*" In [9] the dependability concept is further discussed and the dependability concept is described as containing the following attributes:

- availability, i.e. readiness for service
- reliability, i.e. continuity of service
- safety, i.e. absence of severe adverse consequences on user or environment
- integrity, i.e. absence of improper system alterations
- maintainability, i.e. ability to permit repairs or modifications

Closely related to dependability is vulnerability. The concept of vulnerability is troublesome to define, since the definitions used vary widely depending on context and policy. In many of the definitions the phrase "*susceptibility to injury or attack"* can be found in some form. The concept is used to characterize a system's lack of resilience with respect to different threats that may cause loss of availability, integrity or confidentiality. These threats may originate within as well as outside the system of interest. The vulnerability can be manifested in the system of interest as well as in the supporting or enabling systems.

Related to the concept of vulnerability is the concept of *threat*. A threat may be a natural disaster or other natural phenomenon, a failure of a system element, a faulty system state, an unintentional act by an individual, or a malicious act by an individual or a group. When a threat is present, attempting to exploit vulnerability, this is often called an *attack*.

4 VULNERABILITY AND RISK ANALYSIS

In the analysis of risk and vulnerability of a system, the concept of threat is very useful. It can be seen as a generalization of the concept of hazard, or risk source. Within the field of computer security, many papers discuss the concepts of threat and vulnerability. In the paper [10] is referred to the Common Criteria (ISO 15408) Profiling Knowledge Base (CC-PKB). In the general threat category there are 30 threats, presented in table 1. They can be used for the generation of ideas in a vulnerability analysis of different e-maintenance system functions. Not all of them are applicable to all functions, and the list should not be considered as complete.

Table 1. CC-PKB general threats

1. Administrative errors of commission	16. Malicious code exploitation
2. Administrative errors of omission	17. Unexpected disruption of system or component power
3. Hostile administrator modification of user or system data	18. Recipient denies receiving information
4. Administrator violates user privacy policy	19. Sender denies receiving information
5. A critical system component fails	20. A participient denies receiving information
6. Software containing security-related flaws	21. Legitimate system services are spoofed
7. Failure of a distributed system component	22. Hostile user acts cause confidentiality breaches
8. Hacker undetected system access	23. User abuses authorization to collect data
9. Hacker attempts resource denial of service	24. User errors cause confidentiality breaches
10. Hacker eavesdrops on user data communications	25. User error makes data inaccessible
11. Cryptanalysis for theft of information	26. User errors cause integrity breaches
12. Hacker masquerading as a legitimate user or as system process	27. User errors undermine the system's security features
13. Message content modification	28. User's misuse causes denial of service
14. Exploitation of vulnerabilities in the physical environment of the system	29. User abuses authorization to modify data
15. Social engineering	30. User abuses authorization to send data

A system can be attacked by threats in many ways. In the description of an accident or other adverse event as a combination of events and conditions, the different threats can be described as in Table 2.

Table 2. Threats related to different parts of an accident description

Causes	There are many threats that can directly affect the system or system element under study.
Barriers (probability reduction)	Some threats may attack the barriers, safeguards or safety functions, intended to reduce the probability of an incident.
Incidents	The incident is any event, resulting from an attack.
Barriers (mitigation)	In order to mitigate the impacts of an incident, barriers, safeguards or safety functions are often introduced. They, too, may be subject to threats.
Impacts	An attack may have impact directly on the system under study, on the persons involved and on the physical and business environments.

An important difference between commonly performed risk analyses and vulnerability analyses is how barriers or safeguards are taken into account. In risk analyses they are normally assumed to be active while in the vulnerability analyses they are assumed to be subject to threats as well.

Among the risk or vulnerability analysis methods available, special attention should be given to the well-known FMEA (Failure Mode and Effects Analysis) [11], HAZOP (Hazard and Operability Study) [12] and ETA (Event Tree Analysis) [13]. They all propagate from cause to effect, from threat to consequence, taking the available barriers or safety functions into account. However, there is no need for establishing a brand new methodology, with its own acronym. Instead we can use the possibility of combining a number of tools, chosen according to our knowledge and experience. For vulnerability assessment, one possible set of tools or supporting methodologies consists of the tabular Failure Mode and Effects Analysis (FMEA) sheet, the elicitation of requirements/functions of the system under study, the list (possibly augmented) of general threat categories, elements of barrier analysis, and Event Tree Analysis (ETA). These tools or supporting methodologies have to be combined into a methodology which is systematic, easy to document and verify.

Table 3. Tools and supporting methodologies for the proposed vulnerability analysis approach

FMEA sheet	The FMEA sheet is useful for the documentation of the knowledge gained during the vulnerability analysis process. It should include only the appropriate columns. The analysis is preferably restricted to a qualitative one, leaving out the estimation of probabilities and impact values.
Elicitation of requirements and functions	When performing of risk and vulnerability analyses, the system knowledge is of utmost importance. The knowlege representation can be formalized, making use of a function-malfunction analysis [15]. For the system under study, as well as its elements, a comprehensive list of requirements – in particular the required functions – should be compiled.
General threat categories list	The general threat categories list in Table 1 is one of several possible guides for the identification of possible threats and attacks. It should be used as a set of guide-words (cf the HAZOP methodology [12]) to apply to the different requirements/functions.
Barrier analysis	Threats can be eliminated, attacks obstructed, and the impact of incidents reduced by different kinds of barriers or safety functions. [16], [13]
ETA	The impacts of a postulated incident may differ very much according to which barriers are active and also according to other conditions. The various scenarios are possible to describe, using Event Trees, one separate tree originating from each of the identified incidents.

Experiences from earlier practical work and more recent research [14] indicate that a combination of the afore-mentioned tools and supporting methodologies provides a useful methodology for the analysis of vulnerability of e-maintenance systems and also of the assets to be maintained. The proposed methodology can be schematically described by Figures 1 and 2 below.

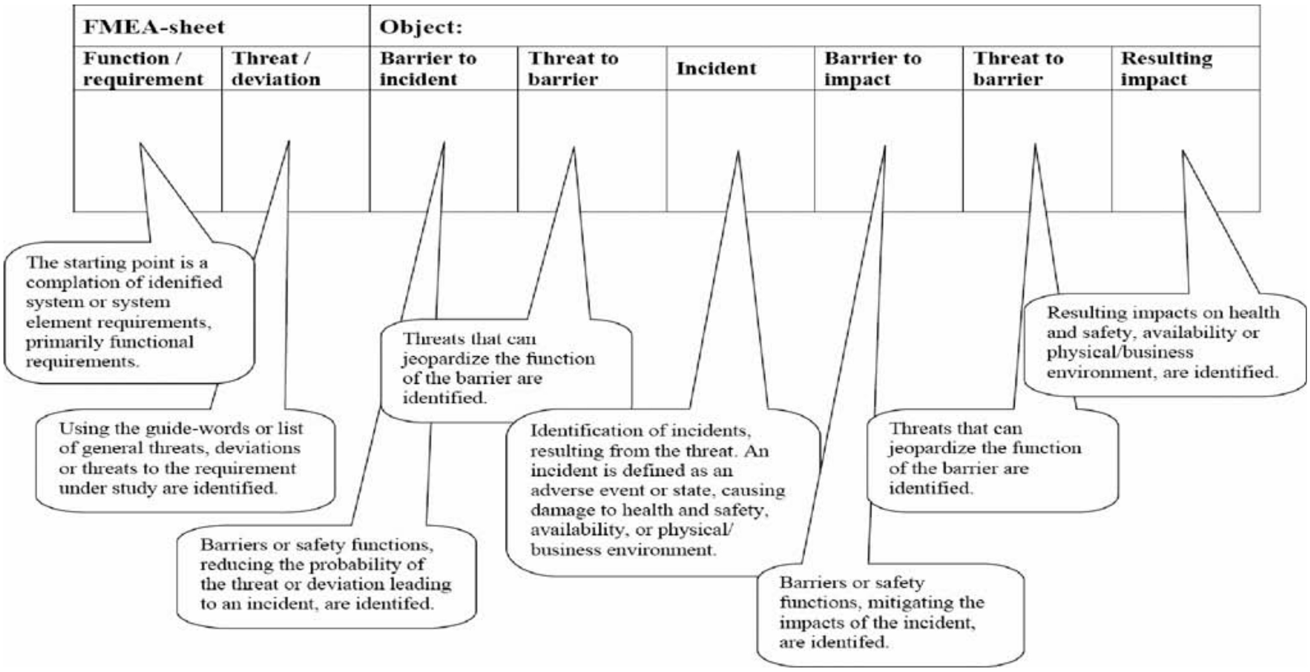


Figure 1. The proposed vulnerabilty analysis methodology

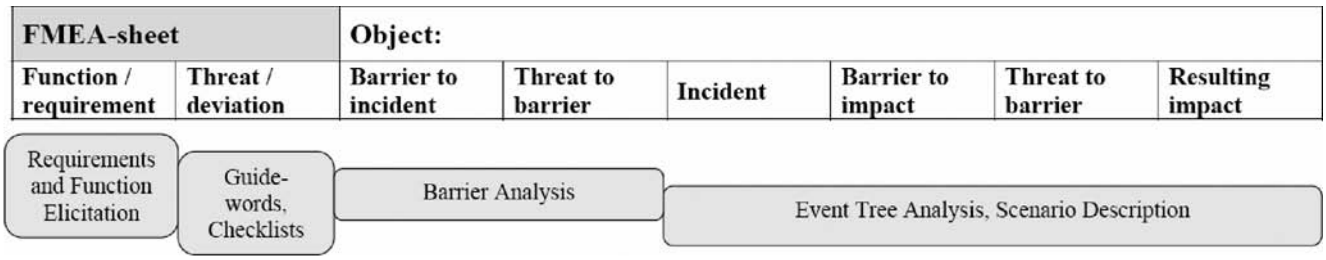


Figure 2. Supporting tools to the proposed methodology

5 CONCLUSIONS

Modern e-maintenance systems, based on the use of information and communication technology, and related assets are subject to many different kinds of threats. The vulnerabilities of the e-maintenance systems, as well as the maintained assets, can be analysed using of existing methodologies and tools. These vulnerabilies should be addressed at an early stage, in the elicitation of requirements on the e-maintenance system. The paper presents one approach, combining a number of well-known tools.

6 REFERENCES

[1] Lee, J. (1998). Teleservice engineering in manufacturing: challenges and opportunities. *International Journal of Machine Tools & Manufacture*, 38, 901-910.

[2] Mathieu, V. (2001). Product services: from a service supporting the product to a service supporting the client. *Journal of Business and Industrial Marketing*, 16(1), 39-61.

[3] Kumar, R., & Kumar, U. (2004). A conceptual framework for the development of a service delivery strategy for industrial systems and products. *Journal of Business and Industrial Marketing*, 19(5), 310-319.

[4] Koç, M., Ni, J., Lee, J., & Bandyopadhyay, P. (2005). Introduction to e-Manufacturing. In: R. Zurawski (Ed.) *Industrial Information Technology Handbook*, Boca Raton: CRC Press, Chapter 97.

[5] McCrossin, R. (2003). Managing risk in telemedicine. *Journal of Telemedicine and Telecare*, 9(Suppl.1), 36-39.

[6] Tettero, O., Out, D.J., Franken, H.M. & Schot, J. (1997). Information security embedded in the design of telematics systems. *Computers and Security*, 16(2), 145-164.

[7] Koc, M. & Lee, J. (2001). A system framework for next-generation e-maintenance systems. Center for Intelligent Maintenance Systems, Univ. of Wisconsin-Milwaukee.

[8] IEV, *International Electrotechnical Vocabulary*, (International Electrotechnical Commission, Geneva, 2006). (Available IEC January 2006) http://std.iec.ch/iec60050

[9] Avizienis, A., Laprie, J.-C., Randell, B. & Landwehr, C. (2002). Basic concepts and taxonomy of dependable and secure computing. *IEEE Transactions on Dependable and Secure Computing*, 1(1), 11-33

[10] Stephenson, P. (2004). Applying impact and vulnerability analysis to risk management. *Computer Fraud and Security*, Issue 2, 16-20.

[11] Stamatis, D.H. (1995). *Failure Mode and Effect Analysis: FMEA from Theory to Execution*. Milwaukee: ASQ.

[12] IEC 61882 (2001). *Hazard and operability studies (HAZOP studies) – Application guide, International Standard IEC 61882.* Genève: International Electrical Commission.

[13] Debray, B., et al. (2004). Defining safety functions and safety barriers from fault and event trees analysis of major industrial hazards. In: C. Spitzer et al. (Eds.), *Proceedings of PSAM7-EREL '04*. London: Springer, pp.358-363.

[14] Söderholm, P. & Akersten, P.A. (2005). Scenario analysis in testability design, using event tree analysis. Paper, submitted for publication.

[15] Iung, B. (2003). From remote maintenance to MAS-based e-maintenance of an industrial process, *Journal of Intelligent Manufacturing*, 14 59-82.

[16] Hollnagel, E. (2004). *Barriers and Accident Prevention, or how to improve safety by understanding the nature of accidents rather than finding their causes*. Aldershot: Ashgate.

AE CONDITION MONITORING: CHALLENGES AND OPPORTUNITIES

JZ Sikorska[a] & D. Mba b

[a] *Imes Group Ltd, Central Ave, Rockingham WA, Australia.*

[b] *Department of Automotive, Mechanical and Structural Engineering,*
Cranfield University, Bedfordshire, United Kingdom.

Abstract: For those entering an engineering field for the first time, be it as a research student or industrial practitioner, it becomes quickly apparent that the plethora of academic and popular literature on a subject never paints a complete picture. This is especially true when a field is relatively new. Such is the case for acoustic emission monitoring, particularly when applied to incipient fault detection in rotating machinery. This imbalance causes the uninitiated to perceive AE as the holy grail of incipient fault detection. Conversely, those who in the past tried and failed, condemn AE forever. Whilst newcomers to research or consulting repeat the mistakes of the past as they struggle to make a contribution. This paper moves to restore the balance by discussing some of the problems facing acoustic emission users, as well as suggesting ways forward to overcome present challenges and obstacles.

Key Words: Acoustic emission, condition monitoring, rotating machinery, incipient failure detection, cavitation, slow-speed bearings

1 INTRODUCTION

Success in an academic environment generally depends on having a prolific publication record, particularly in well-recognised journals. This has two relevant consequences to this paper's discussion. Firstly, results need to be published quickly and often. Consequently, many research conclusions are based on very small datasets. The more realistic the experimental setup, the longer meaningful results take to acquire, so data is most often collected from simplified laboratory rigs. Good quality journals carefully scrutinise the submitted work, but in a new field, finding suitably qualified reviewers is difficult. Secondly, data and conclusions are often skewed towards "the good news", whilst contradictory or ambiguous results are explained away or simply ignored. After all, it is basic human nature to promote one's successes and hide one's failures. Positive research outcomes uphold the individual's reputation, promote his or her institution and encourage further funding. Therefore, researchers are naturally reticent to publicize failed hypotheses. Negative or ambiguous results if submitted for publication, are also less likely to be accepted. Unless the work comprehensively and positively disproves a hypothesis, publishers and readers alike can disregard the work with "if only they'd…the outcome may have been different". The consequence is continuously repeated mistakes, slowing the field's overall progress.

Similarly, in the industrial sector, those introducing a new technique or expanding an existing field are often financially dependant on maintaining their intellectual advantage. Mystery equals margin, so to speak. After all, if everyone were to have the same level of understanding, by definition, there would be no experts; therefore no consultants, no patents and no reason to invest in leading edge technology. Therefore, it is in best interests of these experts to keep their knowledge to themselves. Although many publish in popular industry journals (which are rarely peer-reviewed), the content of these articles is heavily skewed towards what they do, rather than how they do it. Again, problems and weaknesses are unlikely to be discussed unless a unique solution is being proposed. Consequently, those starting out take longer to learn the basics (which typically involves significant trial and error). Newcomers are reticent to publicly discuss their epiphanies; more often than not, these revelations are not new to the field, just new to an individual.

A comprehensive review of the AE applied to a range of rotating machines was recently presented in [1] and it is not the intent of this paper to repeat this exercise. Instead, this paper highlights some of the challenges and obstacles facing those attempting to detect incipient faults in machinery components with AE. Unless otherwise referenced, problems and complications are based on the authors' experiences. For the sake of brevity, it is impractical to display the data supporting each of these observations; they are presented to initiate discussion and to raise awareness that certain challenges exist.

2 ACOUSTIC EMISSION OVERVIEW

2.1 Collecting AE signals

Acoustic Emission (AE) is defined as the range of phenomena associated with structure-borne and fluid-borne propagating waves generated by the rapid release of energy from localised sources within and/or on the surface of a material. Typical sources of acoustic emissions include plastic deformation, microfracture, wear, bubble collapse, friction and impacts. When these waves reach the surface of the material they can be detected and measured by acoustic emission sensors (typically piezoelectric devices). Acoustic emission signals are generally only a few microvolts to millivolts in amplitude when measured at the sensing element and range from several kHz to a few MHz. At these frequencies, signals are strongly attenuated in air, so a suitable couplant is required between the sensor and material to ensure signal transmission.

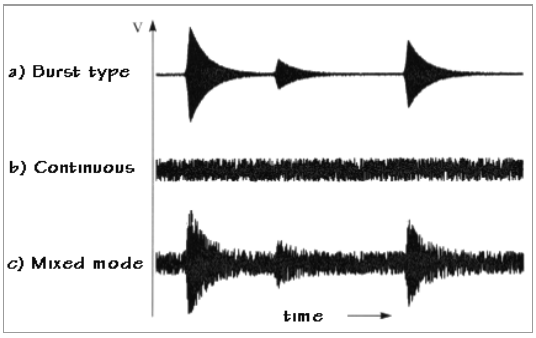


Figure 1: Types of AE signals.

AE pre-processing involves amplification and filtering to refine the bandwidth and avoid aliasing. Signals are characterised as continuous, burst or mixed mode (See Figures 1-2). Burst emissions are typically discrete transients with relatively short decay times and even shorter rise times. Continuous emissions are bursts that occur too closely together to differentiate between individual events, appearing as an increase in the background signal level. They typically have no distinguishing features other than their amplitude and frequency content. As the name suggests, mixed mode AE contains a number of large individual bursts above a background level of continuous emissions. As most AE is broadband, processing is usually done in the time domain.

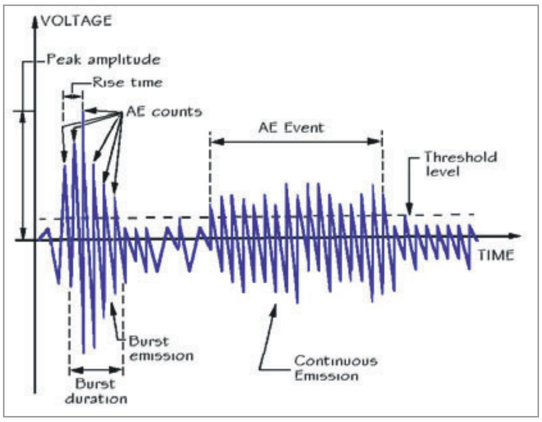


Figure 2: Traditional features of an AE signal.

Hits and events are important concepts to traditional AE signal analysis. A hit is defined as an AE burst that exceeds a certain voltage threshold. Generally speaking, an event occurs when the peak voltage remains above the threshold for consecutive hits. Both hit and event data features are therefore a function of the type and value of the selected threshold. Thresholds are referred to as *fixed*, in which case they are set to an absolute value for the duration of the test (eg. $40dB_{AE}$) or *floating*, where the threshold level is set as a defined amount (a fixed voltage, or fixed number of standard deviations) above the background level. Fixed thresholds are typically used when monitoring static equipment, whilst rotating machinery requires floating thresholds to avoid swamping acquisition hardware when fluctuations in operating conditions cause the background signal level to rise. Unfortunately, the ability of COTS (custom off the shelf) AE hardware to manipulate thresholds is limited. Modern AE systems are theoretically capable of detecting and analysing up to 20,000 hits per second.

Whilst hit features are extracted from a sensor's amplified filtered voltage output, continuous signal descriptors are generally extracted from the signal envelope or from filtered, digitized waveforms. The number of simultaneous waveforms that can be collected depends on the amount of data, which is a function of acquisition rate and resolution, and the acquisition system's digital bandwidth: 4 channels of 16-bit waveform data, collected at 5MHz, will result in 40 megabytes (MB) of data per second. This data must pass across the computer's PCI bus, which currently has a bandwidth of 127MB per second, to the hard disk for storage. Hard disks are typically restricted to writing data at approximately 40MB per second. As a result, undertaking multi-channel AE data acquisition consumes a large proportion of the computing power available in standard desktops and portable computers. It is therefore difficult to perform real-time advanced signal analysis of digitized waveforms, such as discrete-wavelet or short-time-frequency transformation, on unoptimized consumer-grade hardware.

2.2 Processing AE signals

As AE signals travel through a material, they are attenuated, dispersed, reflected and refracted. By comparison with stationary structures, rotating machines are structurally very complex, making interpretation of signals and diagnosis of faults even more challenging than when testing stationary plant.

Depending on the dominant failure mechanism, AE signals collected from rotating and reciprocating machines can be burst, continuous or mixed. Degradation is generally marked by a change in the signal type (eg. continuous to burst), or one or more of its signal features (eg. decrease or increase in burst amplitude). As no two acoustic emission signals are ever the same, most AE processing is strictly statistical. Features, extracted in either hardware or software, are trended over time. Experts peruse the trends for changes and make deductions about the underlying conditions. Due to the signal changes resulting from an AE signal's transmission path, it is particularly important that AE sensors be placed close to potentially faulty components and in the same location from test to test.

Both continuous and burst AE signals from rotating and reciprocating can be highly non-stationary. This further complicates fault identification when using traditional AE approaches. Fortunately, recent work by the first author has shown that this non-stationarity can be managed by grouping data relating to a particular known condition into ensembles and calculating averages for each group. Variability in resulting trends is quantified by calculating confidence limits. Changes in the underlying fault conditions are observed as changes in the mean parameter, and/or its confidence limits [3]. Unfortunately, this process is difficult to do with traditional AE hardware and software.

Neural networks can also be used to find patterns and correlations with the data, and commercial software is available. However, making sense of neural network outputs is not as straightforward as the supplier's brochures might make it seem.

3 GENERIC AE ISSUES

3.1 Signal to Noise Ratio

The most quoted benefit of AE monitoring in the literature is its apparent high signal-to-noise ratio (SNR). This claim is presumably based on its high frequency range when compared to typical mechanical background noise of rotating plant (less than 20kHz). It is true that noise the audible frequency range has limited affect on AE measurement particularly when viewing signals in the frequency domain. However noise refers to more than just extraneous sound. It is defined as any unwanted part of the data set and therefore to ensure valid analysis conclusions, needs to be eliminated (or quantified) at the time of data collection or during post-processing. A variety of techniques have been developed over the past 20 years to denoise condition-monitoring data, with varying degrees of success when applied to AE signals. This is, in part, because what constitutes noise and what is regarded as genuine data changes from one application to the next; thus it is impossible to develop a single set of processing techniques that will be applicable in all situations. As a result, publications relating to noise management tend to be application specific; unfortunately, most relate to traditional AE applications such as crack monitoring.

When collecting acoustic emission data from rotating machinery, noise can originate from some of the following external sources:

(a) Residual, steady background noise from electronic components;

(b) Aperiodic transient events, such as valve activity or electro-magnetic interference (EMI) or radio frequency interference (RFI);

(c) Quasi-periodic transients events that occur approximately the same interval apart, such as rubbing of seals or bearing faults (when not looking for incipient seal or bearing faults);

(d) Low frequency modulation of the background continuous AE level, also common in reciprocating machinery signals or due to misalignment;

(e) Genuine AE from mechanisms not of interest to the analysis being performed.

Separating the wheat from the chaff is problematic and considerable time will be spent analysing data that may in fact have no bearing to the conditions being analysed. Although good acquisition practices can reduce the risk of noise infiltration, there is no way to eliminate it entirely.

Acoustic emission signals, because they are small and weak, are very sensitive to noise ingress. As frequencies get higher, amplitudes tend to decrease and vulnerability to noise contamination increases. Furthermore, certain types of noise only become problematic at higher frequencies. AE signals from rotating machines are even more at risk because they are often collected using broadband non-resonant sensors measuring signals up to 1-2MHz. These are less sensitive that resonant sensors used for structural AE work. Consequently, noise identification, management and mitigation needs to be very carefully considered.

Problems are compounded when machinery is driven by variable frequency drives (VFD). These control the speed of a motor by changing the voltage frequency. Unfortunately, they emit pulses of broadband, high frequency electro-magnetic radiation. This couples to the equipment or AE instrumentation resulting in very large, broadband, impulsive noise bursts superimposed onto AE signals at a switching frequency of 10-500kHz. Noise bursts appear very similar to genuine acoustic emissions, particularly when viewed in isolation (i.e. hit data) and are captured by AE instrumentation as if they were genuine AE hits (See Figure 3).

Furthermore, hydro-mechanical machines such as centrifugal pumps have a number of genuine mechanisms that generate acoustic emissions. Some of these ultimately lead to failure, whilst others are of little consequence and can probably be ignored. Therefore an important part of any fault identification process is separating wanted from unwanted signals. The following key issues relevant to AE machinery monitoring have been recognised or identified [3]:

(a) Current AE hardware does not provide sufficient noise immunity against interference from variable frequency drives. This cannot be removed during post-processing as signals have very similar characteristics to genuine AE events. Therefore, instrumentation source power must be electromagnetically isolated from the VFD.

(b) Electro-magnetic interference (EMI) and radio frequency interference (RFI) often remain after frequency filtering due their very large amplitudes.

(c) Averaging can be used to detect certain noise artefacts in AE signals, such as those associated with exceeding the frequency-amplitude characteristics of on-board amplifiers (a type of clipping). However, a large number of samples per average are required otherwise inadequate resolution in the frequency domain hides frequency peaks.

(d) When analysing continuous AE signals, waveforms must be much longer than individual discrete bursts, so the statistical properties are accurately calculated.

(e) Denoising techniques, such as wavelets (eg. Haar) for waveform data and Swansong filtering for hit features, are very effective at dividing mixed-mode acoustic emission signals into discrete and continuous parts. These can be quantified separately, improving fault diagnosis.

Although adaptive filtering and neural networks have also been applied for segregating specific signals into noise-data components, this paper's authors have not been able to replicate the results for generic signals.

3.2 Hardware Options

In some ways AE hardware is following an evolutionary path well worn by other data acquisition technologies. Traditionally, AE analysis was performed using semi-proprietary systems with 5 or 6 figure price tags. These were primarily designed for structural monitoring and therefore have significant capacity to acquire, characterise and store information pertaining to discrete AE events (i.e. hits). Unfortunately, their capability to capture high-fidelity, high-resolution continuous and quasi-continuous AE signals over an extended duration is severely limited. This was because cheap computing technology was simply unable to process signals quickly enough. As predicted by Moore's Law (which states that the complexity of integrated circuits, the basis of a computer's processing power, doubles every 2 years [4]), over past 10 years data throughput capabilities of desktop PCs have increased a thousand fold and current variants are able to manage several streams of AE waveform data simultaneously. Therefore, it is now possible to continuously monitor AE activity for all speed and load conditions. Outputs can also be feed into DCS (4-20mA) systems [5, 6].

Although handheld "AE" acquisition hardware has been around for some time (in the guise of SPM, SEEPens, Holroyd Memo etc) their functionality was very specific and relevant to particular applications. All were proprietary and as reported in [7] none of these are interchangeable or directly comparable. Fortunately, these technologies are now maturing and more generically capable options are slowly becoming available.

On the horizon, we start to see the next evolution: smart sensors that understand what they are looking at and report their understanding on the state of the equipment, rather than streaming raw data to a host computer for analysis. This facilitates a much more comprehensive and widespread application of AE monitoring. Unfortunately, squeezing a multi-megasample signal processor into a matchbox, running it from batteries and returning the data, is not a simple challenge. Especially when faced with a duopoly of AE equipment manufacturers with a significant investment in traditionally styled hardware. Nevertheless, a Perth company is currently developing this new form of AE technology. Their system involves developing smart, self powered, sensors that report back AE information and its implications over low speed, wireless and internet links. Signal processing requirements are uploaded to the sensor for each application, allowing their use in a wide variety of applications. Targeted at ad-hoc and remote installations, this technology will allow the implementation of scalable AE systems that will integrate into existing SCADA and CBM software.

3.3 Processing Options

Most investigators apply AE technology for one-shot analysis and/or time-based trending, and rely heavily on measurements being collected under the same process conditions. This approach can be fraught with errors as AE varies significantly with most operating variables relevant to the machine (eg. flow, load, temperature, pressure etc). On variable speed and load machines it is particularly difficult to take consistent results. Yet even with consistent duplication of test conditions, one-off testing is not suitable for fast-developing faults, or intermittent problems inexorably intertwined with the machine's operating conditions (eg. seal degradation due to intermittent flashing of the fluid). In these situations, continuous monitoring and real-time (or near-real) processing is the only viable option. Unfortunately, COTS AE systems are poorly suited to this task, particularly if waveforms are to be collected (rather than RMS) for advanced analysis. Fortunately, it is not impossible [5, 6].

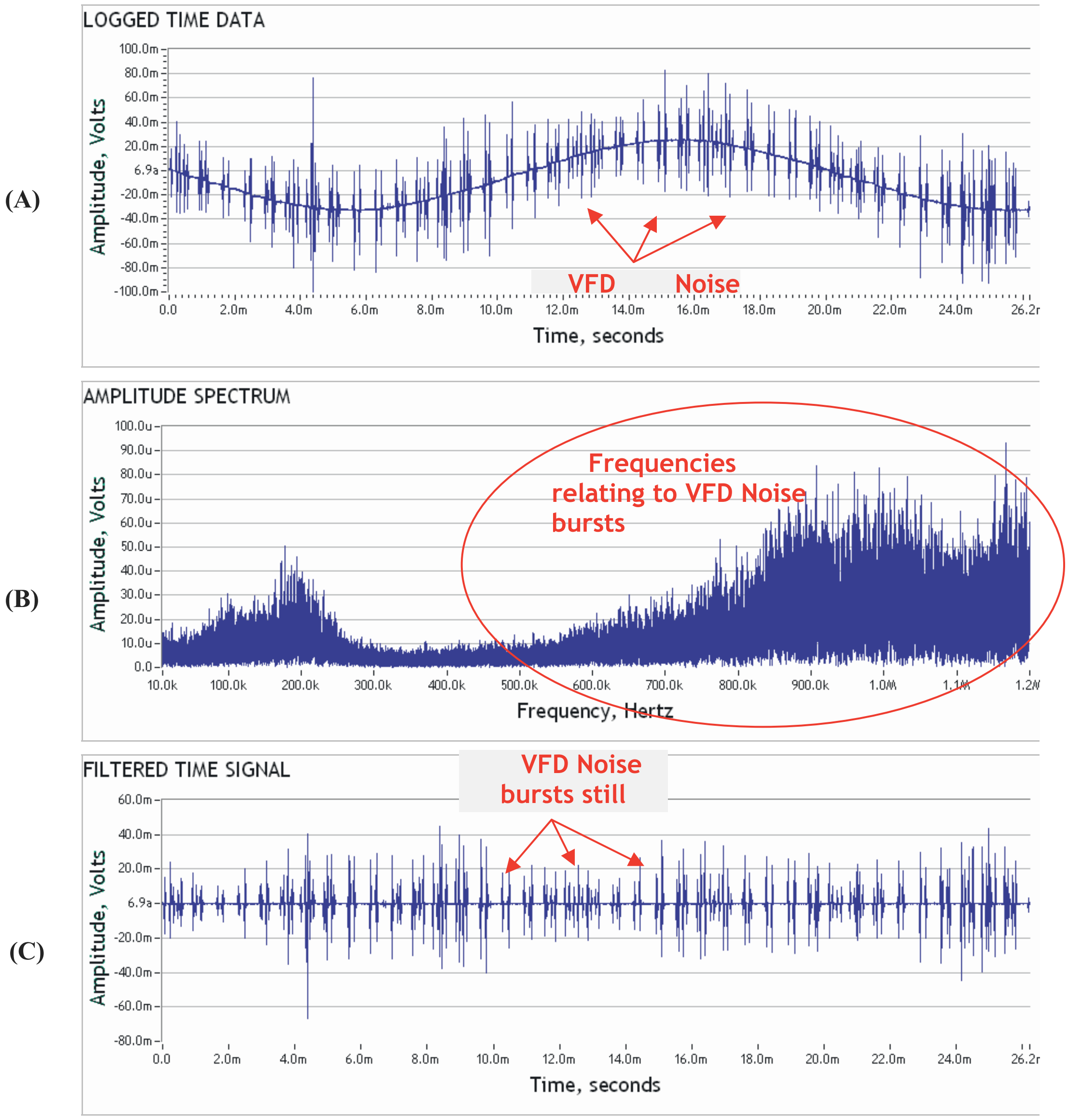


Figure 3: Effect of frequency and wavelet filters on pump AE signal corrupted by a 100Hz frequency sinusoidal signal and high frequency impulsive noise, latter emanating from a VFD. (a) Original time signal, (b) FFT, and (c) 10kHz –1000kHz bandpass filtered.

Traditional AE software packages were designed for structural AE work; their analysis and graphing capabilities are inadequate for condition monitoring. Not only is difficult to suitably digest the data (eg. ensemble averages, confidence limits), time-based graphing functionality is often limited. Each day of continuous monitoring can result in gigabytes of raw data. This volume of information cannot be analysed manually. Very quickly, even accompanying process data and setup conditions become difficult to track. To overcome the limitations of COTS software for continuous applications, a data-driven management system was presented in [5]. The approach is based on a relational database, and has no inherent functional limitations on number of channels, tests or post-processing routines. The number of concurrent users (analysing data) depends primarily on the database management software. Millions of data points can be analysed as desired, with auditable rigour, and minimal manual intervention.

4 APPLICATION SPECIFIC ISSUES

4.1 Rolling Element Bearings

Hertzian contact stress, induced by the relative rolling action of the bearing elements in contact with inner and outer races, and rubbing between damaged mating surfaces within the bearing generate a number of different failure modes. These include flaking, brinelling, fluting, spalling, pitting and seizures. Whilst vibration monitoring, the traditional technique for rolling element bearing fault detection, measures the response of the structure to the developing fault (eg. a change is surface roughness, clearance reduction, developing spall), AE sensors measure the actual degradation mechanism itself.

Therefore it is not surprising that literature extensively reports AE as offering superior early detection capabilities over vibration analysis for bearing faults [8-14]. At low bearing speeds (less than 100rpm), where vibration energy transmitted is small, AE is particularly effective [15-18]. An increase in hit counts is the most commonly reported change with developing outer race defects although the peak amplitude of periodic bursts (at the outer-race frequency) has also been interrogated successfully. Interestingly, although most work rules in favour of AE over vibration for early detection, some contrary reports have been published, particularly when monitoring higher speed bearings [16, 19]. This may be because collective results show that AE counts are sensitive to the level and grade of lubricant within the bearing [20-22], whilst amplitude if affected by speed. Hit features are also dependent on the threshold selected and operating speed (the higher the bearing frequency, the higher the hit rate) [8, 23]. Collectively, these problems make it particularly difficult to identify the origins of a defect using AE hits, although a variant of the standard count parameter has been successfully employed for bearing diagnosis [8]. For most applications however, a multi-feature approach is probably required to successful diagnose outer-race defects using AE.

Due to the sensitivity of AE to surface roughness, recent work has found that defect lengths can be estimated from the AE signal [12, 24]; defects generate AE bursts, the amplitude and duration of which is proportional to defect depth and length respectively (see Figure 4). This information is not obtainable with traditional condition monitoring techniques.

The same overwhelming success has not been reported for inner race defect detection. It has been concluded that emissions from inner race defects are not of sufficient strength/energy to be detectable above the operational background noise [11]. This is probably because of the complex, continuously varying transmission path between the inner race defect and sensors mounted on the bearing casing. The result is an attenuated and amplitude-modulated AE signal (at multiple frequencies rather than just the defect frequency) that complicates diagnosis.

More recently, frequency analysis of AE signals (or demodulated AE) and other advanced techniques have been used to improve defect identification in both inner and outer races. These include autoregressive coefficients, Point Process Spectral Averaging, adaptive noise techniques and pattern recognition analysis. Although, none of these techniques are currently being applied beyond the instigating research institution, this will inevitably change in time.

Proprietary “black box” devices for monitoring bearings, based on a form of AE (called stress wave or high frequency resonance analysis), have been available to industry since the 1970s. To improve detectability, these devices use the transducer’s natural frequency to amplify the signal, later removing this carrier frequency through some form of demodulation. Some systems require a proprietary sensor, whilst others use a generic accelerometer. Systems that necessitate their own sensors claim that the resonant response has been carefully “tuned” to the resonant frequency of the bearing, yet interestingly these frequencies vary from 30 kHz to several hundred kilohertz. Fortunately, impulsive events and lubrication breakdown tend to induce broadband excitations, so will most likely be detected equally well by any of the systems. Although they have been used for some time now, numerous formal and anecdotal reports indicate problems acquiring repeatable and reliable results with these technologies. This is because system outputs often (but not always) depend on equipment load, sensor location and experience of the user. Furthermore, under some fault conditions, background levels increase swamping any increase in impulsive events that form the basis of black box outputs. Unfortunately, not all devices automatically account for changes in operating conditions, requiring the user to interpret the data accordingly. More details on these devices are given in [7]. A recent study reports that traditional AE analysis offers earlier detection of bearing faults than shock pulse (a stress wave variant) monitoring, both of which offering better detection that vibration analysis [13].

Finally, it should be noted that most results on the applicability of AE to bearing monitoring have been performed under laboratory test conditions designed to reduce the effects of background noise. In industrial applications signals will be more complex and in some circumstances advanced denoising techniques may be required. However as [25] describes, adequate indication of incipient failure is still possible in the presence of noise.

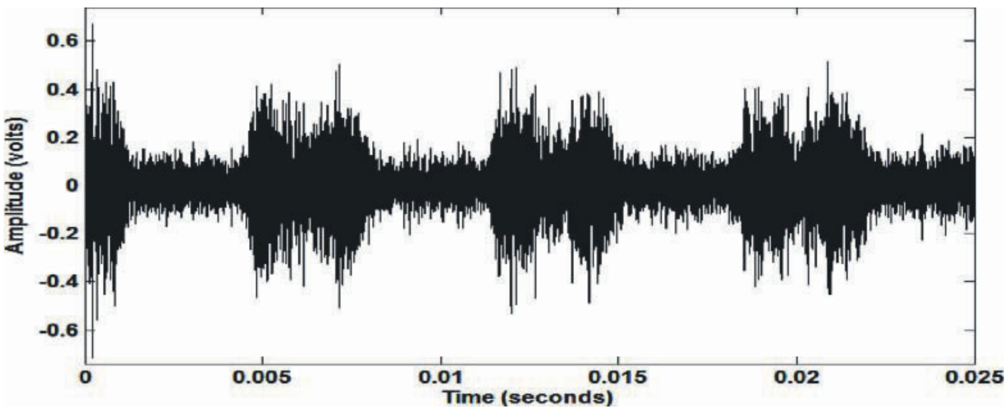


Figure 4: Duration of AE bursts is indicative of the outer race defect length [16]

4.2 Mechanical Seals and Journal Bearings

Lubrication and wear is a major problem for many machinery components, particularly bearings and seals. Collectively, work relating AE signals to wear seems comprehensive. Relationships have been established between many AE parameters and wear variables, including type of wear, rate and volume [26-32], although not all researchers have equal success quantifying all variables. AE has also been shown as very sensitive indicator of lubrication conditions and changes in lubricant properties [21, 22, 28].

Industrial studies have successfully shown that AE is an effective technique for detecting incipient failures of mechanical seals [33-37] with fluctuations in AE energy features relating to instability in the lubricating film. Leakage, dry running and cavitation in the seal gap have also be detected using acoustic emission [33, 35]. Some researchers also determined that acoustic emission variations coincide with torque and temperature variations [34]. Unfortunately, a number of contradictory reports have been published. Presently, the development of a robust detection and prognosis system for mechanical seals is being thwarted by difficulties in generating failures in a controlled, but realistic, environment [3, 38, 39].

The high sensitivity of AE to wear also leads to potential difficulties in interpretation. It is widely reported that AE signals in wear tests on various component types are significantly higher on start-up whilst the surfaces faces “bed-in” [27, 35]; if a test is paused and restarted, signals generally return to their previous levels [40]. In the authors’ work on mechanical seals however, this “bedding-in” process has been observed occurring repeatedly throughout seal’s life without any cessation of operation, probably due to minor changes in process conditions that cause the faces to adjust their relative positions; when the seal was subsequently removed for inspection, no faults or abnormal wear was detected. On other occasions when the seals were left in service, these increased AE levels eventually dropped [3]. Another complication reported with AE seal monitoring, is that some failure modes result in a decrease in the signal amplitude or energy, contrary to expected trends [31, 34, 35]. This may be due to an increase in the lubricating film causing leakage, but decreasing wear and thus AE levels. Alternatively, in certain ceramic materials (from which many modern mechanical seal faces are fabricated), high temperatures initiate the formation of an oxide layer. When this occurs, AE levels have been observed to decrease, despite an increase in wear rate [31]. As most AE wear studies have been undertaken in laboratories with metal on metal samples, the application of results to ceramic seal faces in industrial machines must be undertaken with caution. Fortunately, one or two studies using ceramic materials have been performed, with positive outcomes [3, 29], so incipient failure detection of modern seal faults is certainly realisable. In a practical environment, consideration regarding the positioning of sensors is also required, as AE transducers located on the seal or pump housings will transmit AE signals generated by the pumpage [3].

Interestingly, although the application of AE monitoring to journal bearing degradation is perhaps the most obvious application for the technology in rotating equipment management, it is also the least investigated. Difficulties arise in locating AE sensors sufficiently close the bearing. Additionally, journal bearings are often used in high-speed machinery, where harmonics of traditional mechanical faults (eg. gear mesh frequency) intersect the AE bandwidth and thus can be difficult to remove. Nevertheless, some successes with journal bearing monitoring have been reported [41, 42].

4.3 Gearboxes

In gearboxes, sources of AE include tooth cracking, gear wear and lubrication breakdown [43-46]. Both crack initiation and propagation has been successfully detected, using a variety of traditional and advanced AE signal analysis, and precedes fault identification with traditional monitoring techniques. Gear pitting and cracking has also been detected on individual teeth [45-47]; AE parameters generally increase with pit size. Again, AE appears capable of faults prior to either vibration or wear debris analysis and various failure modes can be differentiated [48]. Unfortunately, under non-isothermal conditions as present in a normal bearing, energy and amplitude features are also affected by load and speed [40, 49], which may complicate diagnosis in an industrial environment; these features do not change when temperature is kept constant [44]. Diagnosis is also more difficult at extremely high speeds [45] or when testing normalised gears [47]. It has also been reported that AE bursts disappear after a defect is established [50]. A similar effect is commonly reported when monitoring AE levels from rolling element bearings due to the disappearance of impulsive events into an increased background level.

Another major drawback is the difficulty of placing AE sensors close to the region of interest. Attenuation in a gearbox can be significant due to the large number of individual interfaces and potential transmission paths [45]. Consequently, researchers have placed AE sensors on the rotating gear and employed a slip-ring to record activity from the rotating sensor [43, 44]. A comparative analysis of the placement of the senor on the rotating gear and the bearing pedestal showed much greater sensitivity on the rotating gear. Consequently, the development of miniature AE sensors with telemetric capabilities will have a major impact the feasibility of AE fault detection for various types of rotating plant.

4.4 Pump Monitoring

Unfortunately, attempts to detect incipient component faults in centrifugal pumps using acoustic emission testing have been complicated by the sensitivity of AE to its operating state [51]. The event of bubble collapse results in release of AE energy and the application of AE to early stages of cavitation has been well documented [51-56]. Once cavitation progresses, AE levels generally drop [57] due the attenuation of signals by large vapour pockets. This effect is more pronounced in larger pumps. Trends at smaller flows seem less consistent, perhaps due to the combination of recirculation and cavitation that results in more complex bubble dynamics. Others have reported that AE signals from cavitation can be differentiated from signals measured under low-flow recirculation and best-efficiency point conditions [53, 58]. Increases in AE signals at low flow, presumed to be from recirculation, may be more significant than changes at high flows [52]. Researchers disagree whether signals are affected by sensor location on the pump [52-54, 56]. This may be because pumps are not equally prone to recirculation, which is the condition that can probably only be adequately diagnosed by mounting sensors on the suction flange [52]. However, more work is required to confirm this hypothesis.

Although not as widely studied as cavitation detection, flow changes under normal suction conditions have also been detected with AE [52]. Results from pump performance tests showed that, particularly in large pumps where hydraulic changes are significant, various AE energy features can be very effective for detecting movement away from best efficiency flow. Unfortunately, results from smaller pumps have been less conclusive, particularly at low flows, probably due to the relatively small changes in hydraulic energy across the range of flows, and consequent sensitivity to the testing process. However, even in these pumps, consistent patterns in hit energies were observed. This suggests that a generic monitoring technique for all pumps may be possible with AE, although more work is required to verify this hypothesis.

A complication is discriminating between collapsing bubbles at pump surfaces and those occurring in the middle of the cavity (and therefore not likely to cause damage). It is generally accepted that bubbles releasing only small amounts of energy are not problematic. Yet large number of small energy events will cause background AE level to rises swamping any underlying problem. Acoustic emissions from other fault mechanisms can also complicate diagnosis, although the converse is more likely. Fortunately, advanced filtering and signal processing techniques can be applied to extract pertinent information from, and/or uncover hidden trends in the AE signal (see Figure 5).

Another challenge that needs to be managed is the large dynamic range required of measurement instrumentation; signal amplitudes can increase by 60-80dB under cavitation or recirculation, as compared to conditions around best efficiency. Therefore if equipment is selected to detect small changes during normal operating conditions, extreme flows will swamp electronics causing unusual artefacts in the digitized signal, leading to possible misinterpretation of data. Even software has difficulty representing both extremes with adequate resolution. In some large pumps, very small deviations from best efficiency cause signal clipping, which easily goes unnoticed. Baseline values from pump to pump can also vary by 40-60dB, requiring the AE user to hold a variety of sensors and preamplifiers.

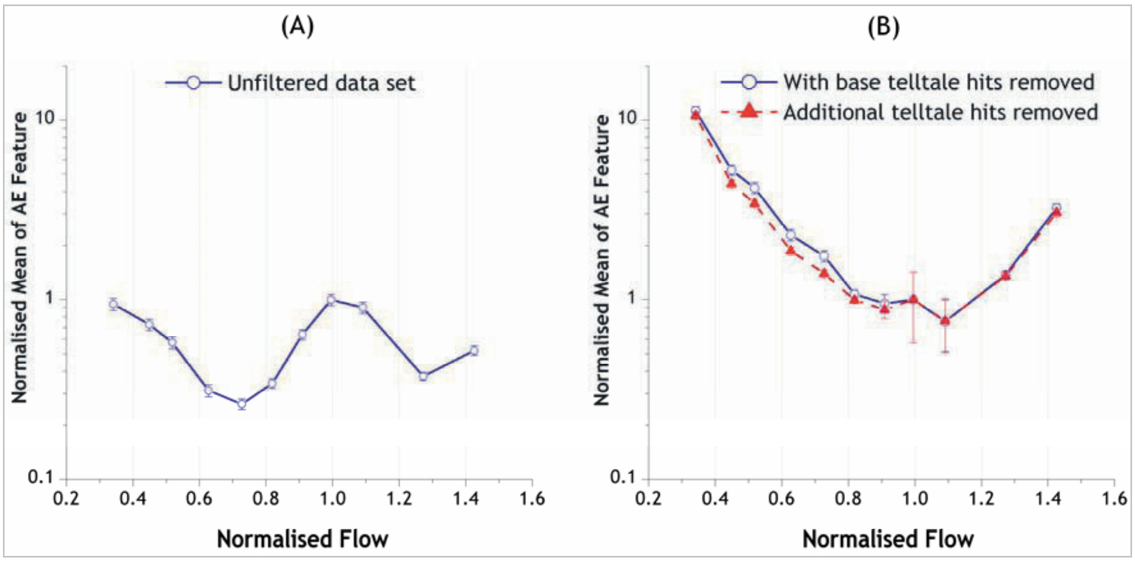


Figure 5: Improved trending is obtained by advanced filtering techniques. (A) Shows an AE feature versus flow without filtering, whilst (B) shows the same feature after filtering. In both cases data has been normalised with respect to the mean feature value at BEP. (Vertical lines indicate 90% confidence limits, which increase because little data remains after filtering.)

4.5 Reciprocating machinery

AE monitoring has been applied successfully to the detection of valve leakage in reciprocating compressors [59] and a hand-held or online device using externally mounted AE sensors is now available to industry. Interestingly, although piston and rider ring wear would seem to lend itself to AE monitoring, no work in this area has been reported. This may change with the advent of cheap, wireless sensors that facilitate mounting on the piston itself.

The application of AE to engine monitoring (both diesel and internal combustion variants) is currently one of the most active research areas. Traditional analysis involved measuring RMS trends and comparing them to known-good conditions. Sources of AE that have been successfully detected include problems with valves or injectors, piston ring with cylinder liner interactions, lubricant flow and/or blowby, as well as the combustion/compression process itself [60-68]. Different failures modes occur at varying times during the engine's cycle, which helps diagnosis [69].

Although the authors have no personal experience with engine AE monitoring, one of biggest issues reported in literature involves discriminating the various AE sources. Thus researchers are actively investigating advanced signal processing techniques and source location algorithms that will better segregate and/or identify faults. Neural networks are being widely applied for detecting and diagnosing faults in diesel engines. In general, acoustic emission is reported as having superior detection capabilities over traditional monitoring technologies for a variety of failure modes. However, neural network classification success is generally improved when a variety of sensor types are used [69]. The outputs of these pattern recognition algorithms also lend themselves to automated faultfinding, which is required prior to widespread industrial application. Detecting fault conditions amidst a high background levels or when faults result in smeared signals, is also challenging. Again, this is currently being tackled with research into the application of advanced signal processing algorithms.

The ability of acoustic emission monitoring over traditional methods for detecting certain will ensure that AE will ultimately be integrated into mainstream engine monitoring systems, however due to the processing challenges, this may take some time.

5 CONCLUSIONS

Although AE may have failed to deliver in the past or may not be living up to current expectations, progress is continuing. With research effort and industrial experience, challenges are being overcome and problems are finding solutions. AE may never the single answer to all monitoring applications, but nor is any other technique. Good engineering is about finding the right tool for the job. In many applications, such as journal bearing wear, slow speed bearing faults, mechanical seal degradation and engine monitoring, viable alternatives simply do not exist. Consequently, acoustic emission analysis will play an increasingly important role.

Relating AE signals to their originating sources is a challenge for all machinery monitoring applications. The AE community is still far from accurately predicting responses to theoretical AE sources on static structures, much less for the complex shapes of typical rotating plant. It is therefore inevitable that problems in interpretation and false diagnoses will occur for some time to come. Error will be reduced when sensors can be placed close to the faults; multiple sensors in some form of virtual array will facilitate mathematical or deductive source location. For this, we need small, cheap, wireless transducers with wide dynamic range and inbuilt analysis functionality. Until then, good engineering judgement, as well as a better dissemination of problems and solutions, is fundamental to ensuring analysis of AE data is robust and reliable.

6 REFERENCES

[1] Mba D. & Rao R.B.K.N. (2006) Development of Acoustic Emission Technology for the Condition Monitoring and Diagnosis of Rotating Machines: Bearings, Pumps, Gearboxes, Engines and Rotating Structures. The Shock and Vibration Digest, 38(1), 3-16.

[2] Holroyd T. (2000) *Acoustic Emission & Ultrasonics*, Oxford, UK: Coxmoor Publishing Company.

[3] Sikorska J.Z. (2006) The Application of Acoustic Emission Monitoring to the Detection of Flow Conditions in Centrifugal Pumps. PhD, University of Western Australia.

[4] Moore G.E. (1965) Cramming More Components onto Integrated Circuits. Electronics Magazine, April.

[5] Sikorska J.Z., Kelly P.J. & Pan J. Development of an AE Data Management and Analysis System. Mechanical Systems and Signal Processing, In Press, Corrected Proof.

[6] Fan Y., Gu F. & Ball A. (2005) Condition Monitoring of Rolling Element Bearings Using Advanced Acoustic Emission Signal Analysis Technique. In Proceedings of the 18th International Congress of Comadem, Cranfield University, 491-498, Cranfield University Press.

[7] Sikorska J.Z. & Pan J. (2003) A Review of Acoustic Emission Techniques for the Diagnosis of Faults in Rotating Machinery. In Proceedings of the ICOMS 2003, Perth Australia, IEAust.

[8] Tan C.C. (1990) Application of Acoustic Emission to the Detection of Bearing Defects. In Proceedings of the Institution of Engineers Australia Tribology Conference, Brisbane, 110-114, IEAust.

[9] Tandon N. & Nakra B.C. (1992) Comparison of Vibration and Acoustic Measurement Techniques for the Condition Monitoring of Rolling Element Bearings. Tribology International, 25(3), 205-212.

[10] Li C.J. & Li S.Y. (1995) Acoustic Emission Analysis for Bearing Condition Monitoring. Wear, 185(1-2), 67-74.

[11] Shiroishi J., Li Y., Liang S., Kurfess T., et al. (1997) Bearing Condition Diagnostics Via Vibration and Acoustic Emission Measurements. Mechanical Systems and Signal Processing, 11(5), 693-705.

[12] Al-Ghamdi A.M. & Mba D. (2006) A Comparative Experimental Study on the Use of Acoustic Emission and Vibration Analysis for Bearing Defect Identification and Estimation of Defect Size. Mechanical Systems and Signal Processing, 20(7), 1537-1571.

[13] Tandon N., Yadava G.S. & Ramakrishna K.M. A Comparison of Some Condition Monitoring Techniques for the Detection of Defect in Induction Motor Ball Bearings. Mechanical Systems and Signal Processing, In Press, Corrected Proof.

[14] Price E., E.D. (2005) Detection of Severe Sliding and Pitting Fatigue Wear Regimes through the Use of Broadband Acoustic Emission. Proceedings of the Institution of Mechanical Engineers. Part J, Journal of engineering tribology, 219(2), 85-98.

[15] Smith J.D. (1982) Vibration Monitoring of Bearings at Low Speeds. Tribology International, 15(3), 139-144.

[16] McFadden P.D. & Smith J.D. (1984) Vibration Monitoring of Rolling Element Bearings by the High Frequency Resonance Technique - a Review. Tribology International, 17(1), 3-10.

[17] Jamaludin N., Mba D. & Bannister R.H. (2001) Condition Monitoring of Slow-Speed Rolling Element Bearings Using Stress Waves. Proceedings of the IMechE Part E: Journal of Process Mechanical Engineering, 215, 245-271.

[18] Mba D., Bannister R.H. & Findlay G.E. (1999) Condition Monitoring of Low-Speed Rotating Machinery Using Stress Waves. Part 1. Proceedings of the I MECH E Part E Journal of Process Mechanical Engineering, 213, 153-170.

[19] Williams T., Ribadeneira X., Billington S. & Kurfess T. (2001) Rolling Element Bearing Diagnostics in Run-to-Failure Testing. Mechanical Systems and Signal Processing, 15(5), 979-993.

[20] Miettinen J. & Salmanperä P. (2000) Acoustic Emission Monitoring of Grease Lubricated Rolling Bearings. In Proceedings of the Comadem 2000, USA, 21-30, MFPT Society.

[21] Miettinen J. & Andersson P. (2000) Acoustic Emission of Rolling Bearings Lubricated with Contaminated Grease. Tribology International, 33, 777-787.

[22] Jamaludin N., Mba D. & Bannister R.H. (2002) Monitoring the Lubricant Condition in a Low-Speed Rolling Element Bearing Using High-Frequency Stress Waves. Proceedings of the IMechE Part E: Journal of Process Mechanical Engineering, 216, 73-88.

[23] Huguet S., Godin N., Gaertner R., Salmon L., et al. (2002) Use of Acoustic Emission to Identify Damage Modes in Glass Fibre Reinforced Polyester. Composites Science and Technology, 62(10-11), 1433-1444.

[24] Al-Ghamdi A.M., Cole P., Such R. & Mba D. (2004) Estimation of Bearing Defect Size with Acoustic Emission. Insight, 46, 758-761.

[25] Morhain A. & Mba D. (2003) Bearing Defect Diagnosis and Acoustic Emission. Proceedings of the IMechE Part J Journal of Engineering Tribology, 217, 257-272.

[26] Mechefske C.K., Sun G. & Sheasby J. (2002) Using Acoustic Emission to Monitor Sliding Wear. INSIGHT, 44(8), 490-497.

[27] Boness R.J., McBride S.L. & Sobczyk M. (1990) Wear Studies Using Acoustic Emission Techniques. Tribology International, 23(5), 291-295.

[28] Boness R.J. & McBride S.L. (1991) Adhesive and Abrasive Wear Studies Using Acoustic Emission Techniques. Wear, 149, 41-53.

[29] Benabdallah H.S. & Boness R.J. (1999) Tribological Behaviour and Acoustic Emissions of Alumina, Silicon Nitride and SAE52100 Uder Dry Sliding. Journal of Materials Science, 34, 4995-5004.

[30] Matsuoka K., Forrest D. & Tse M.-K. (1993) On-Line Wear Monitoring Using Acoustic Emission. Wear, 162-164, 605-610.

[31] Hisakado T., Suda H. & Sekine M. (1998) Relation between Lubricated Wear of Ceramics and Acoustic Emission Characteristics. Tribology Transactions, 41(2), 209-216.

[32] Dvoracek J., Petras J. & Pazdera L. (2000) The Phase of Contact Damage and Its Description by Help of Acoustic Emission. Journal of Acoustic Emission, 18, 81-86.

[33] Bloch H.P. (1978) Acoustic Incipient Failure Detection Systems Are Successful at Exxon. The Oil and Gas Journal, Feb. 6, 66-72.

[34] Kataoka T., Yamashina C. & Komatsu M. (1987) Development of an Incipient Failure Detection Technique for Mechanical Seals. In Proceedings of the Fourth international pump symposium, 121-129.

[35] Miettinen J. & Siekkinen V. (1995) Case Study: Acoustic Emission in Monitoring Sliding Contact Behaviour. Wear, 181-183, 897-900.

[36] Ferguson I., Neill G., Gill J.S., Brown E.R., et al. (1998) Condition Monitoring of Rotating Seals Using Acoustic Emission. In Proceedings of the 23rd European Conference on Acoustic Emission Testing, Vienna.

[37] Mba D., Roberts T., Taheri E. & Roddis A. Application of Acoustic Emission Technology for Detecting the Onset and Duration of Contact in Liquid Lubricated Mechanical Seals. INSIGHT, Under review.

[38] Flepp B. (1999) Wear Diagnosis of Mechanical Seals with Neural Networks. PhD, Swiss Federal Institue of Technology.

[39] Neill G. (1998) Pc Based Diagnostic System for the Condition Monitoring of Rotating Machines. PhD, Heriot-Watt University.

[40] Toutountzakis T. & Mba D. (2003) Observations of Acoustic Emission Activity During Gear Defect Diagnosis. NDT & E International, 36, 471-477.

[41] Sato I., Yoneyam T., Sato K., Tanka T., et al. (1991) Applications of Acoustic Emission Techniques for Diagnosis of Large Rotating Machinery and Mass Production Products. Acoustic Emission: Current Practice and Future Directions, ASTM STP 1077, 287-304.

[42] Sturm A. & Uhlemann D.-I.S. (1985) Diagnosis of Plain Bearings by Acoustic Emission Analysis. Measurement, 3(4), 185-191.

[43] Tan C., Irving P. & Mba D. (2005) Diagnostics and Prognostics with Acoustic Emission, Vibration and Spectrometric Oil Analysis for Spur Gears; a Comparative Study. INSIGHT, 47(8).

[44] Tan C.K., Irving P. & Mba D. A Comparative Experimental Study on the Diagnostic and Prognostic Capabilities of Acoustics Emission, Vibration and Spectrometric Oil Analysis for Spur Gears. Mechanical Systems and Signal Processing, In Press, Corrected Proof.

[45] Singh A., Houser D.R. & Vijayakar S. (1999) Detecting Gear Tooth Breakage Using Acoustic Emission. Journal of Mechanical Design, 121, 587-593.

[46] Robinson J.C. (2001) Detection and Severity Assessment of Faults in Gearboxes from Stress Wave Capture and Analysis. In Proceedings of the Society for Machinery Failure Prevention Technology (MFPT) 2001 Annual Conference.

[47] Miyachika K., Oda S. & Koide T. (1995) Acoustic Emission in Bending Fatigue Process of Spur Gear Teeth. Journal of Acoustic Emission, 13(1/2), S47-S53.

[48] Soires E. & Negro A.A. (1997) Condition Monitoring of a Gear Box Using Acoustic Emission Testing. Materials Evaluation, 55(2), 183-187.

[49] Dornfeld D. & Handy C. (1987) Slip Detection Using Acoustic Emission Signal Analysis. In Proceedings of the 1987 IEEE International Conference on Robotics and Automation, 1868-1875.

[50] Raad A., Zhang F., Randall B. & Sidahmed M. (2003) On the Comparison of the Use of AE and Vibration Analysis for Early Gear Fault Detection. In Proceedings of the 8th Western Pacific Acoustics Conference, Melbourne Australia.

[51] Hutton P. (1969) Detecting Acoustic Emission in the Presence of Hydraulic Noise. Nondestructive Testing, 2(2), 111-115.

[52] Sikorska J.Z. & Hodkiewicz M. (2005) Comparison of Acoustic Emission, Vibration and Dynamic Pressure Measurements for Detecting Change in Flow Conditions on a Centrifugal Pump. In Proceedings of the COMADEM, Cranfield University, Comadem.

[53] Alfayez L. & Mba D. (2005) Detection of Incipient Cavitation and Determination of the Best Efficiency Point for Centrifugal Pumps Using Acoustic Emission. IMechE, Part E: Journal of Mechanical Process Engineering., 219(4), 327-344. *Under Review.*

[54] Neill G.D., Reuben R.L., Sandford P.M., Brown E.R., et al. (1997) Detection of Incipient Cavitation in Pumps Using Acoustic Emission. In Proceedings of the IMechE, Part E: Journal of Process Mechanical Engineering, 267-277.

[55] Derakhshan O., Houghton J.R., Jones R.K. & March P.A. (1991) Cavitation Monitoring of Hydroturbines with RMS Acoustic Emission Measurement. ASTM STP 1077, 305-315.

[56] Darling J. & Johnston D.N. (1991) The Use of Acoustic Emission for Condition Monitoring in High-Pressure Hydraulic Pumps. In Proceedings of the Eurotech Direct 91, Birmingham UK, 137-143, IMechE.

[57] Alfayez L., Mba D. & Dyson G. (2005) The Application of Acoustic Emission for Detecting Incipient Cavitation and the Best Efficiency Point of a 60kw Centrifugal Pump; Case Study. NDT&E. *In press.*

[58] Neill G.D., Brown E.R., Reuben R.L., Sandford P.M., et al. (1998) Detection of Recirculation in Pumps Using Acoustic Emission. In Proceedings of the COMADEM, Melbourne, Australia, 651-660.

[59] Gill J.D., Brown E.R., Twite M., Horner G., et al. (1998) Monitoring of a Large Reciprocating Compressor. In Proceedings of the COMADEM'98, Monash University, Australia, 317-326.

[60] Azzoni P., Moro D. & Rizzoni G. (1998) Time-Frequency Signal Analysis of the Acoustic Emission of Formula 1 Engines. In Proceedings of the IEEE, 441-444.

[61] Fog T.L., Hansen L.K., Larsen J., Hansen H.S., et al. (1999) On Condition Monitoring of Exhaust Valves in Marine Diesel Engines. In Proceedings of the Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop, 554-563.

[62] Sharkey A.J.C., Chandroth G.O. & Sharkey N.E. (2000) Acoustic Emission, Cylinder Pressure and Vibration: A Multisensor Approach to Robust Fault Diagnosis. In Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN 2000), 223-228.

[63] Wu J.-D. & Chuang C.-Q. (2005) Fault Diagnosis of Internal Combustion Engines Using Visual Dot Patterns of Acoustic and Vibration Signals. NDT & E International, 38(8), 605-614.

[64] El-Ghamry M., Steel J.A., Reuben R.L. & Fog T.L. (2005) Indirect Measurement of Cylinder Pressure from Diesel Engines Using Acoustic Emission. Mechanical Systems and Signal Processing, 19(4), 751-765.

[65] Douglas R.M., Steel J.A. & Reuben R.L. A Study of the Tribological Behaviour of Piston Ring/Cylinder Liner Interaction in Diesel Engines Using Acoustic Emission. Tribology International, In Press, Corrected Proof.

[66] Nivesrangsan P., Steel J.A. & Reuben R.L. (2005) AE Mapping of Engines for Spatially Located Time Series. Mechanical Systems and Signal Processing, 19(5), 1034-1054.

[67] Pontoppidan N.H., Sigurdsson S. & Larsen J. (2005) Condition Monitoring with Mean Field Independent Components Analysis. Mechanical Systems and Signal Processing, 19(6), 1337-1347.

[68] Nivesrangsan P., Steel J.A. & Reuben R.L. Acoustic Emission Mapping of Diesel Engines for Spatially Located Time Series--Part Ii: Spatial Reconstitution. Mechanical Systems and Signal Processing, In Press, Corrected Proof.

[69] El-Ghamry M., Reuben R.L. & Steel J.A. (2003) The Development of Automated Pattern Recognition and Statistical Feature Isolation Techniques for the Diagnosis of Reciprocating Machinery Faults Using Acoustic Emission. Mechanical Systems and Signal Processing, 17(4), 805-823.

A FRAMEWORK TO ASSESS DATA QUALITY FOR RELIABILITY VARIABLES

M.Hodkiewicz[1] P.Kelly[2] J.Sikorska[3] & L.Gouws[4]

[1]. University of Western Australia [2]. CASWA Pty Ltd [3]. IMES Group Ltd [4]. Melikon Pty Ltd

Abstract: This paper presents a framework for assessing the impact of the data collection process on the validity of key measures in reliability. The quality of data is affected by many organisational and behavioural factors. The aims of developing this framework are to (1) identify inputs/steps that have the most significant impact on the quality of key performance indicators such as MTTF (mean time to failure) and MTTR (mean time to repair), (2) identify 'weak' links in the data collection process, and (3) identify potential remedial actions. Development of this framework will assist the understanding of assumptions used in reliability calculations and improve the quality of underlying data and the data collection process. Consequently, this is a vital step in the continued development and use of data based decision-making models for reliability assessment.

Key Words: Data collection, Key performance indicators, reliability assessment

1. BACKGROUND

Engineering Asset Managers rely on data to support decisions. There has been little published work on how to assess the quality of maintenance data and its fit-for-purpose to support decisions. This problem is compounded by ambiguity over who is responsible for assuring data quality and appropriate use of the data.

The quality of the data is dependent on various attributes including timeliness, accuracy, relevance, completeness and accessibility [1]. There is currently no widely accepted methodology for (i) assessing the quality of maintenance-related data, (ii) confirming that data is fit-for-purpose or (iii) identifying and managing changes to the data collection, storage and use system. Many of the typical problems affecting data quality are discussed in the case-study based work of [2] and the experiences of [3] in the establishing of the OREDA project.

Previous work has identified three key elements required for data quality: knowing-what, knowing-how, and knowing-why. It also suggests that there are three roles associated with the data process: data collectors, custodians, and consumers [4]. This research work also showed that data quality is highly dependent on data collectors knowing-why.

For equipment maintenance and reliability data, the data collector is predominantly a maintenance or production technician. The custodians are the IT support staff responsible for managing various enterprise databases and systems, whilst the main data consumers are (i) reliability engineers who use the data during the determination of long-term maintenance strategies, and (ii) maintenance engineers and maintenance supervisors who use the data when addressing day-to-day maintenance issues.

The framework described herein emphasises the importance of communication between data collectors and consumers, not only about the use of the data but also the implication of poor quality data on the business process.

2. THE 8 STEP PLAN

A framework for evaluating data requirements is depicted in Figure 1.

Step 1: Identify the business need

Quality data is simply defined as "data that is fit for purpose" [4]. Therefore it must, by definition, be assessed in the context of a specific business need (its purpose). After all, multiple consumers of a particular set of data may use it for different purposes and therefore data that is of sufficient quality for one user may not be appropriate for another. Consequently, before any data quality improvement process can be initiated, there must be a clear understanding of the business decisions that must be supported by the data. We describe this as a context-oriented approach to data quality.

Business needs will generally be defined using words such as "measure", "quantify", "improve", "reduce" or "control". To control and improve a business process, its status first needs to be determined and methods to measure the effect of changes then determined. The word "measure" and its various synonyms imply an inherent data requirement.

Step 2: Identify the metrics that will meet the business need

Once a business need has been identified it is possible to determine what metrics can be used to meet that need. Some of the most common Reliability-Availability-Maintainability (RAM) metrics include mean time to failure (MTTF), mean time between failure (MTBF), mean time to repair (MTTR), availability, reliability, logistical delay time and a variety of maintenance related costs.

In addition to identifying what must be measured, it is also important to identify the level of accuracy required and the frequency at which the metrics are to be recalculated (eg. once a month, once a day, 1000 times per second etc). Again, these are context specific issues.

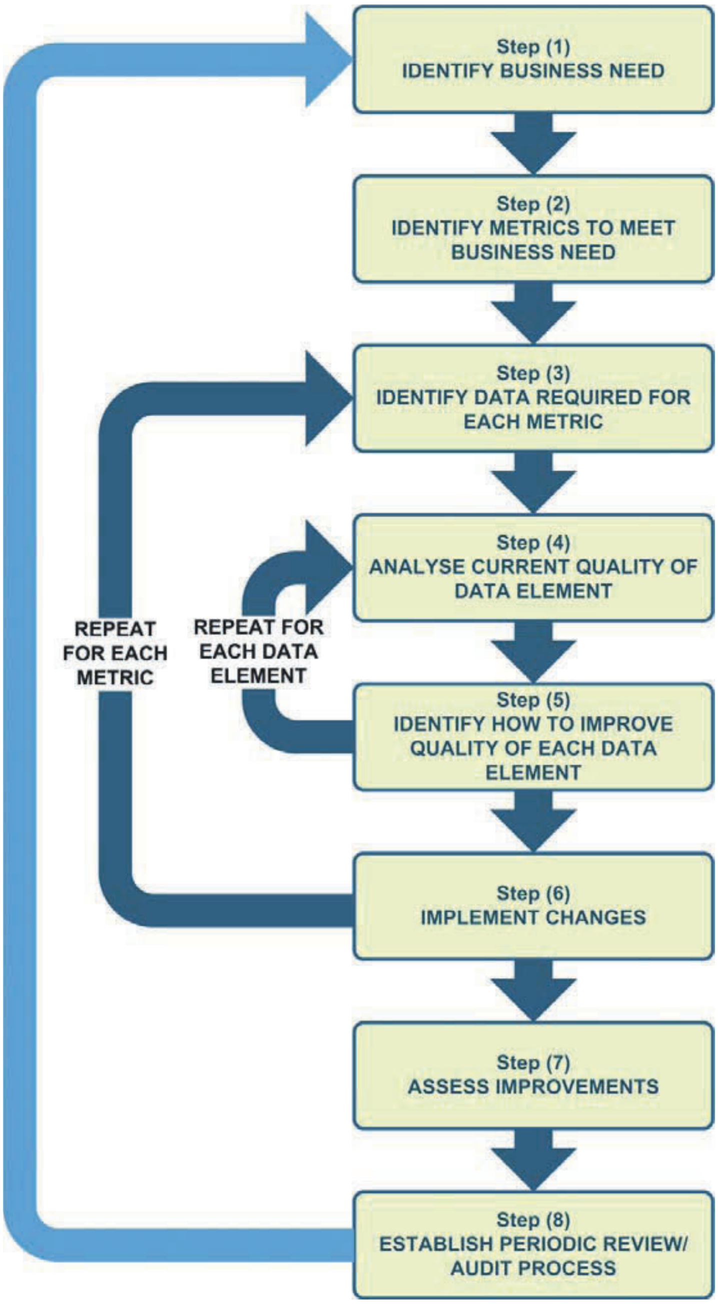


Figure 1: Framework assessing and improving maintenance data quality

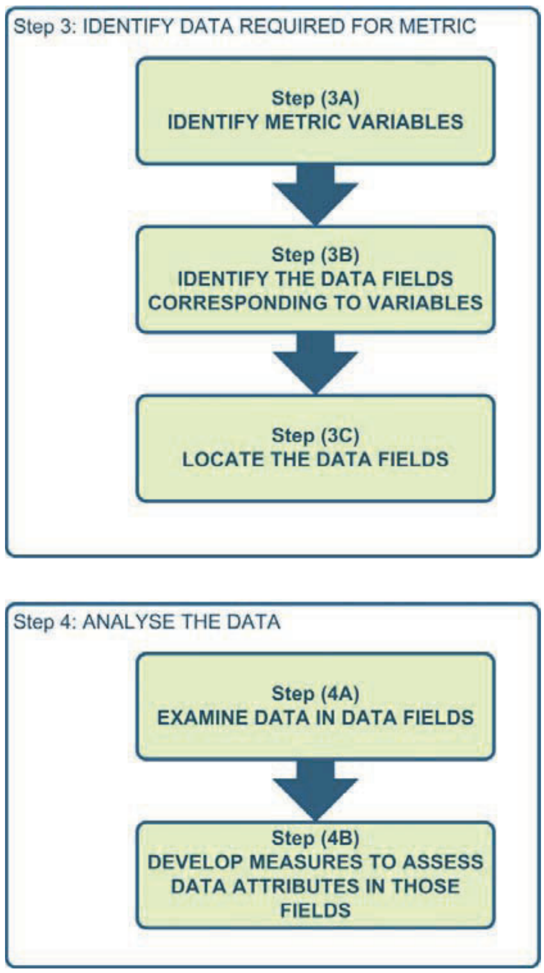


Figure 2: Breakdown of Step 3 and Step 4

Step 3: Identify the data required for each metric

Having selected the metrics, the data required to produce them to the required level of accuracy can be determined. In practice, this actually involves three steps, as illustrated in Figure 2.

A: *Identify metric variables*. By implication, metrics involve some changing variables or dependencies. It is therefore important to identify what the variables are, and how often they need to be acquired.

B: *Identify the data fields that correspond to variables*. Does the variable correspond to a single data field in a database or single gauge to be inspected, or must it be deduced from a number of sources? As the number of data sources increases, so does the risk of error, especially when data is collected manually.

C: *Determine where and how the data is stored.* Data required for RAM analyses is stored in a variety of formats and locations, including equipment databases, computerized maintenance management systems (CMMS) and/or associated ERP systems, spreadsheets, paper records and/or the memories of maintenance and operations personnel. In this paper, we will collectively refer various software systems as Asset Management Databases (AMD).

Step 4: Analyze Data

Once the data has been sourced, its context-specific quality can be assessed. Again, this step can be divided into two smaller steps, shown in Figure 2.

Data sources first need to be manually examined and the current contents qualitatively assessed. *Are the fields full or empty? Is a field codified or free-text? Do the fields contain data relevant to the problem?* This overview influences the development and/or selection of more specific questions for more comprehensive quantitative assessment.

It is suggested that context-specific questions be formulated to collectively determine the following:

Is the data accurate? (Intrinsic data quality)

Is the data appropriate to the business need? (Contextual data quality)

Is the data represented appropriately? (Representational data quality)

Is the data accessible but secure? (Accessibility data quality)

The second step is to develop measures to assess data attributes for the data fields. Required attributes for each of these categories are given in Table 1. The structure and nature of the questions should be dependent on the data variable under investigation and its context. Wherever possible, questions should be designed to elicit a quantitative response. Answers can then be viewed individually and/or as weighted sums, resulting in an overall score of data quality for that particular field.

Table 1: A conceptual framework for data quality (Adapted from [1])

Data quality category	Attributes	Explanation
Intrinsic (Does the data have any inherent quality in its own right?)	Believability	Real and credible
	Accuracy	Accurate, correct, reliable, errors can be easily identified
	Objectivity	Unbiased and objective
	Reputation	Reputation of the data source and data
Contextual (Is the data appropriate to the business need?)	Relevancy	Applicable to task at hand, usable
	Timeliness	Age of the data is appropriate to the task at hand
	Completeness	Breadth, depth and scope of information contained in data
	Appropriate amount of data	Quantity and volume of available data is appropriate
	Value-added	Data gives a competitive edge, adds value to the operation
Representational (Is the data represented appropriately?)	Interpretability	Data are in appropriate language and units and data definitions are clear
	Ease of understanding	Easily understood, clear, readable
	Representational consistency	Consistency formatted and represented, data are compatible with previous format
	Concise representation	Concise, well organized, appropriate format of the data
Accessibility (Is the data accessible but secure?)	Accessibility	Accessible, retrievable, speed of access
	Access security	Access to data can be restricted, data is secure.

Step 5: Identify how to improve the quality of each data element

It is likely that respondents will give very different answers to the data assessment questions posed in the previous step. Ultimately, it will be up to the data consumer to decide whether a piece of data is “fit for purpose”.

Where the data is not fit for purpose, the first step is to meet with the collectors of the data to understand why this is so. One of the greatest reasons for erroneous data is a lack of understanding or appreciation by data collectors on the quality requirements for a piece of data.

Step 6: Implement changes

Unfortunately, a data quality audit will probably result in the identification of more problems than can reasonably be resolved within available timeframes and/or budgets. Therefore, it is suggested that improvements be prioritised based on the complexity of the change required, and the aggregated error caused by data in its present form on the metric and business need(s). A Pareto chart is a good tool to help visualise the changes and decide where time and money should be invested.

Step 7: Assess improvements

By revisiting *Step 3* and using the same measures of data quality, it is possible to express the success or failure of *Step 6* quantitatively.

This (and the next) step require a long-term focus as there may be a significant delay required to allow a meaningful amount of data to accumulate after changes have been made. One of the benefits from *Step 7* is the confidence of having hard data to support “the cause” of improving data quality in the organization.

Step 8: Establish periodic review/order process to verify how well business need is being met.

By periodically reviewing the metrics required to meet the current business needs the relevance of the data is maintained. Realistically, if *Step 7* is to be considered difficult and rarely implemented, then this stage acquires a mythical status. The reality however, is that business requirements change for non-engineering reasons. For example, a decision is made to run the plant at minimum cost, or to ensure availability for a critical period. Or maybe a new CEO is appointed. As mentioned in *Step 5*, the most important people to keep abreast of current business needs and corresponding data requirements are the data collectors; improving their understanding has a profound follow on effect.

3. APPLYING THE 8-STEP PLAN TO ASSET MANAGEMENT DATA

This process described above will be illustrated using one common metric, *Mean Time to Failure*.

Step 1: Define the business need

For the purposes of illustrating the process we are going to assume that the following business need has been defined by a refinery's management team:

"To measure and improve reliability of rolling element bearings in refinery centrifugal pumps."

Step 2: Identify the metrics

The selected metric is: Mean time to failure (MTTF) of bearings in pumps

Step 3: Identify data required

MTTF= Total number of bearing operating hours / Total number of bearing failures

In a perfect system from which only MTTF would ever be calculated, each bearing failure event would have (a) a single unique designator classifying its failure completely (i.e. pump bearing failure), and (b) the age of the bearing at the time of failure. Unfortunately, data storage systems do not always record this information directly. Instead, failure classification or operating hours needs to be deduced from a number of fields, as shown in Figure 3. If age at failure (operating hours) is not available, the date of failure and date of installation are often used instead. Data requirements for this bearing example include the functional location (or equipment number), failure mode (i.e. bearing failure), failure type, event date/time and operating profile. Operating profile is helpful because it can be used to explain extraordinary events (i.e. data outliers) that may not reasonably represent the life of bearings at the refinery, but can significantly skew results.

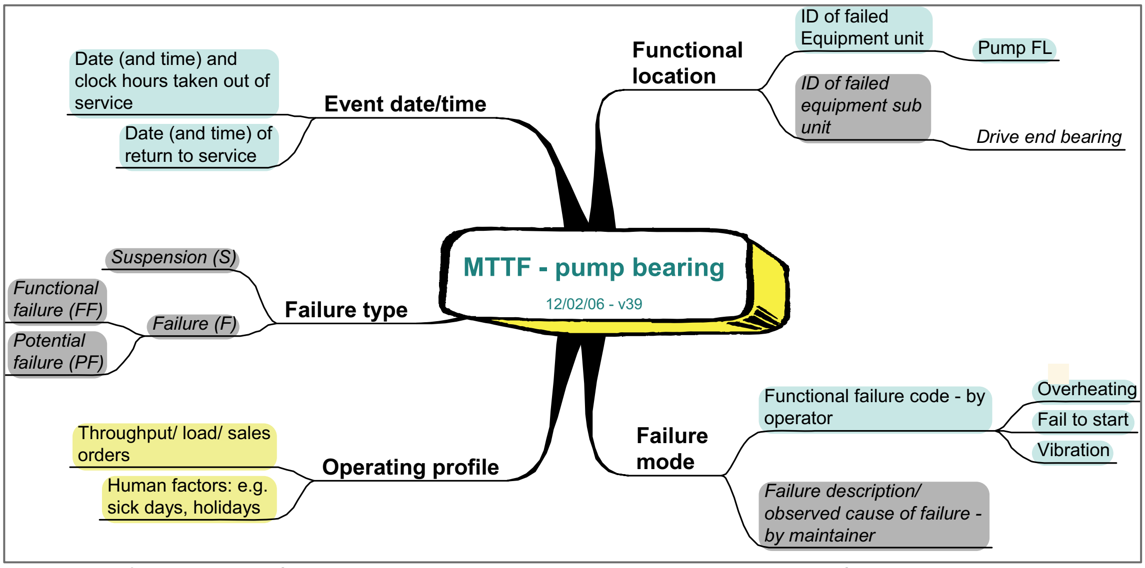


Figure 3: Example of the data required to calculate the MTTF metric for a pump bearing

In Figure 3, the font format (normal/italic) and box colour (grey, yellow, blue) is used to indicate the 'data collector'. In the context of this example,

The operator/ production technician (normal font, blue box) records the date the pump is removed from and returned to service, the functional location of the pump, and the functional failure code. This code identifies 'why' the unit was removed from service.

The maintainer (italic font, grey box) identifies the maintainable item that is the cause of the equipment being removed from service, in this case, the bearing. Further information is then required to describe the failure; example codes are provided in ISO 14224 [5]. Another vital piece of data is to identify if the maintainable item is a *functional failure*, *potential failure* or *suspension*. For this example, a *functional failure* occurs if the bearing has failed causing the pump to shut down. If the bearing is damaged but still functioning at the time of removal, this is described as a *potential failure*. A *suspension* is used to describe the situation when the bearing is replaced in the course of a pump overhaul, when the bearing is still functioning. This may occur if the pump is removed from service to replace a mechanical seal or rebuild the wet-end. The recording of suspensions is vital for the accurate estimation of MTTF. For example, if one pump operating from 1000 hours experiences 2 bearing failures and has a further 2 functioning bearings replaced during replacement of mechanical seals, the true MTTF (accounting for suspensions) is 500 hours. If the suspensions are incorrectly counted as failures, the calculated MTTF is 250 hours. Therefore, omitting the suspension/failure distinction can result in a dramatic under-estimation of the bearing MTTF.

The operating profile (normal font, yellow box) comes from neither the operator nor maintainer. It may include, but is not limited to (1) indicators of the load on the equipment, for example, plant throughput, equipment load, sales orders, (2) the conditions of operation, and (3) who operated the unit. This information is likely to reside in databases outside the CMMS which can be accessed through data warehouses, reports and other IT solutions.

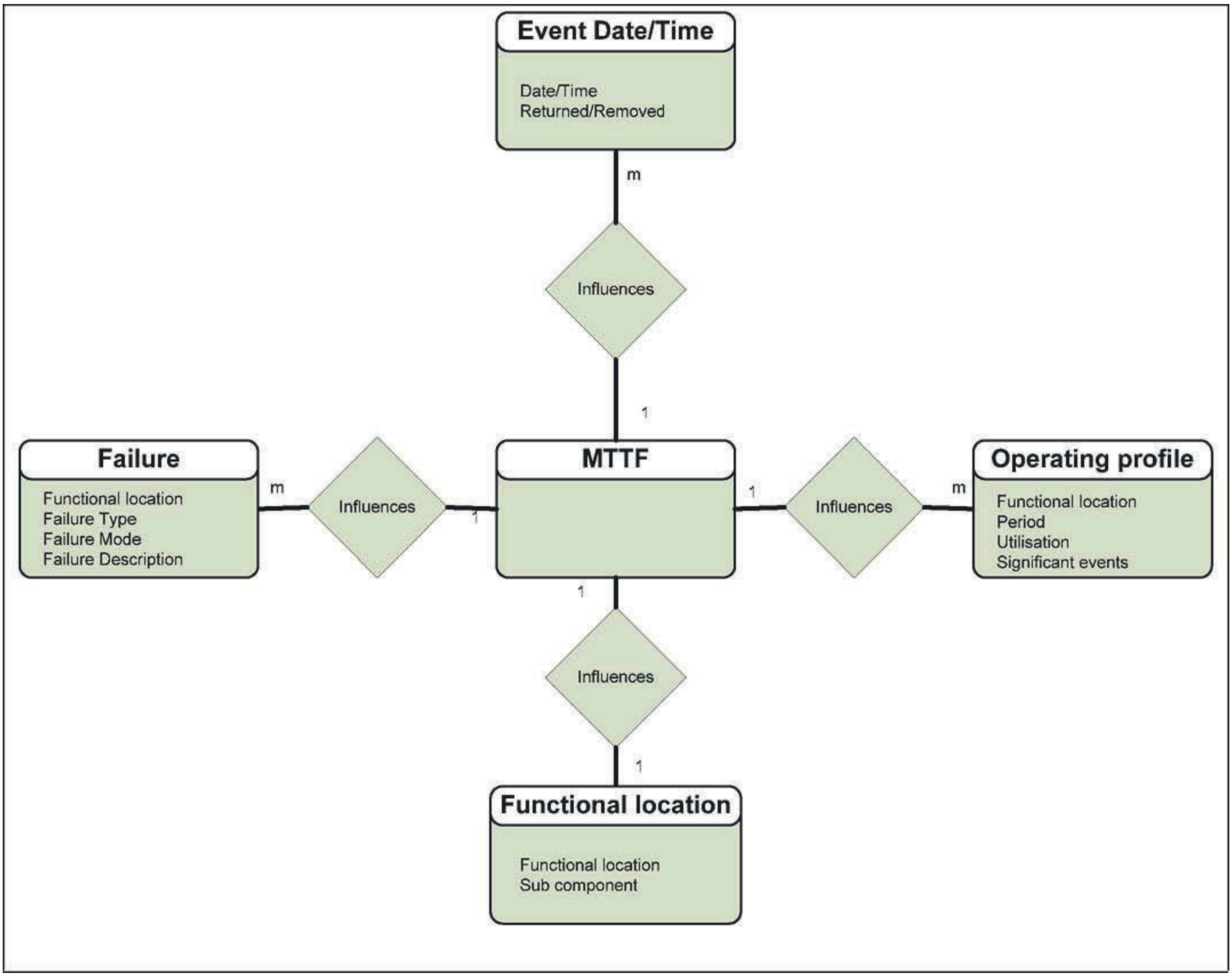


Figure 4: Entity relationship diagram for data shown in the Mind Map in Figure 3

Figure 4 shows the MTTF metric, and its contributing factors from Figure 3, expressed as an Entity Relationship diagram. Database administrators and software developers use ER diagrams to specify the design of data tables and queries. This translation from the data collector's view to the data custodian's without loss of information or form is important as it facilitates accurate communication between data consumers and custodians, this results in better IT systems which result in better quality data.

Step 4: Analyse the data

Step 4a: *Examine data in the data fields*. Investigation of the data may reveal issues such as:

Inaccurate failure date.

Nonsense free-text data (for example "Broken" as a fault description).

Incorrectly selected options.

Prevalence of "other" or "none of the above" selections in restricted entry fields.

Missing data.

Referential integrity problems.

Inappropriate selection of functional location level.

Step 4b: Develop measures to assess data attributes in those fields

In order to measure the data quality of each of the main four fields (functional location, event date, failure mode and failure type) for this example, the authors propose the application of questions about each of the relevant data quality attributes from Figure 1. This is illustrated in Figure 3. For each attribute, one or more questions are required to assess the quality of that attribute in the context of the data field, the Functional location in this example, under investigation. Examples of questions that address each of the data attributes are shown. The questions can be framed to give either yes/no, true/false, scaled (1-5) or numerical (%) replies depending on the data variable under investigation. This illustration is not intended to be comprehensive but illustrative of a general approach.

Table 2: Application of selected data quality attributes to the Functional Location data field

Data quality category	Attributes	Question for the Functional location field in the MTTF calculation	Marking
Intrinsic	Accuracy	The data is correct	Y/N
		The data is accurate	Y/N
	Reputation	I know who collected the data	Y/N
		I know who entered the data	Y/N
		Is the selection of FL affected by who collects the data?	Y/N
		Data in this field is relevant for this analysis	T/F
Contextual	Relevancy	Data in this field is appropriate for this analysis	Y/N
		Data is sufficiently current for our work	Y/N
	Completeness	Data is complete	Y/N
		Are FL data definitions clear to all users (collector, custodian and user)?	Y/N
Representational	Concise representation	Is the information easily retrievable?	Y/N
	Ease of understanding	Are drop-down menus well structured?	Y/N
		If there been changes in the FL structure as the CMMS has been upgraded. Is there a translation between old and new systems?	Y/N
	Representational consistency	Is each FL entered in exactly the same format?	Y/N
		Is the maximum number in any one drop-down menu less than 8?	Y/N
Accessibility	Accessibility	Is the information easily accessible?	Y/N
		Is the information easily obtainable?	Y/N
	Access security	Is there a system to prevent unauthorized changes to the system?	Y/N

Step 5: Identify how to improve the data quality of each element

Based on the results of step 4, we identify that there are some issues with missing data and that some fields are being incorrectly filled in. To address these issues we decide on the following measures:

- Auditing of data at the time of collection.
- Establishing a process for data vetting.
- Careful selection of classifications, Avoid presenting codes.
- Reduce manual data entry. Source date data from other, automated sources (e.g. purchasing system).
- Data entry close to the job in time and space
- Training
- Introduce workflow/process to derive change from exceptions such as selecting "other"

Step 6: Implement changes

The proposed may be ranked in terms of their efficacy and impact. This allows for changes, which have high impact and low to medium cost or difficulty, to be tackled first.

Table 3: Example of table to rank options for improving data quality

Change	Efficacy	Cost/Difficulty
Alter IT systems to check for known garbage values	High	Medium
Two week audit drive on work order data	Medium	Low
Identify data owners and 'goal' them on data quality	High	Very High
Install automated vibration sensors on worst performing pumps	Very High	Medium

Step 7: Assess improvements to data quality

Select a suitable period for review and engage an outside body to apply the metrics of data quality developed in *Step 4* to information collected after the completion of *Step 6*. The resulting independent report will highlight changes that have been effective in improving data quality and assist in identifying where to focus more effort.

Step 8: Establish periodic review/audit process

The executive summaries of all the data quality projects are reviewed annually to assess the relevance of to the current business needs. Operations staff are present at this meeting and feed back their experiences and thoughts on the changes that have been implemented. Where it is clear that the business need has changed, the need to revisit the data quality requirements is discussed. In this case MTTF is a useful base statistic, relevant to a number of business needs. As such improving the quality of its underlying data is likely to deliver value for some time to come.

4. DISCUSSION: GLOBAL ISSUES

Data field links

Figure 4 illustrates how a specific functional location at the unit level is linked to other fields in the CMMS and other databases. The selected equipment component, in this example, a pump bearing, will determine (1) the fields of the sub-unit and maintainable items (2) the equipment-specific corrective maintenance and preventative maintenance activity types, and (3) the equipment-specific failure descriptors and causes. In addition, the functional location may be linked to a serial number and perhaps also a separate equipment identification number (though equipment ID and equipment unit level functional location are often the same number). These numbers are related to installation and design data in the equipment database. There may also be additional links to databases housing reliability centred maintenance analysis or root cause failure analysis information. Careful thought needs to go into both the design and use of functional location structures and the design of links to related fields internal and external to the main database. This mind-map approach to identify all major users of specific data fields coupled with a knowledge of the data required for each of calculated reliability variables can be used to identify (1) key data fields and (2) identify the stakeholders in the data collection, storage and use sequence that need to be consulted if changes are required.

Data collection versus data use

Work by Lee and Strong [4] has shown that data collectors with "why-knowledge" about the data production process contribute to producing better quality data. Overall, improving the knowledge of data collectors is more critical than that of data custodians. *Knowing-why* is gained from experience and understanding of the objectives and cause-effect relationships underlying activities (*knowing-what*) and procedures (*knowing-how*) involved in work processes in organisations [4]. In other words, it is important for the collectors of data required for reliability analysis (the maintainers and operators) to understand why they need to collect the data and how it will be used. If, as is common, the use of the data has not been defined, either copious amounts of data are collected without any defined or proven purpose or, alternatively, insufficient data of any kind is collected. Either situation results in little data of use being available for decision making [6]. The solution is to identify the variables on which key decisions are made, identify the data required for these variables, and communicate both the process and the results to the data collectors. Unless the data collectors appreciate the process and know how, why and by whom the data is being used, they will have no motivation to collect quality data. Furthermore, the more direct and timely the feedback to the data collector about its quality and the direct effect this has on the business need, the more influential this feedback process becomes. Data collectors should know what their data is used for, and given feedback on outcomes following the analysis based on the collected data.

Codes versus text

Many database systems rely on data collectors entering much of their data by making selections from predefined lists, for example, a list of failure mode codes or functional locations. The downside of this is that very little description is supplied to the data collector to help them make their selection, resulting in increased incidence of miscoding or excessive use of exemption codes (e.g. "other"). In extreme cases, normally when using older systems, only the raw codes are presented with explanations given in associated paper manuals or help files. Paradoxically, entering data in a pre-codified form does not make it easier for the computer system to deal with. Choosing a value from a list of options and having the computer generate the code is no harder and represents a very small cost when compared to the subsequent entry and application of inappropriate or incorrect data. Where possible, the authors recommend that data collectors be presented with a rich description of available options. Where there are several similar options available, clarifying text and images should be provided in a context sensitive manner using popup windows or "mouse over" actions. Skilled data entry operators can achieve data entry error rates below 1%[7]. If we charitably describe our operators as being skilled in data entry, that means our 100,000 row maintenance history table may contain 1000 garbage records. The reality is often much more frightening. To pick up on the example presented earlier, screens that allow the input/selection of functional location (equipment/tag/plant ID) are frequently poorly implemented. In the presentation of this data by computer systems, a hierarchical approach is ubiquitous, with much effort and meeting-time devoted to the design of the tree. A problem arises because different users of the system perceive the organization of equipment in different ways. As an example, the electrical system engineer may think of a faulty light globe as '*100W globe in socket 7 of cable run 3 from distribution panel 8*'.Meanwhile an operator sees it as '*100W globe at east end of MCR3 plant west*', and the stores technician as '*100W globe, Supplier 3, electrical consumables*'.

The important point is that the examples in the preceding paragraph are all valid ways of looking at the issue. Ironically, many CMMS only present one hierarchy, thereby disadvantaging users who view the data structure differently. To overcome this problem, a project is currently underway to trial a data-mining portal which allows multiple trees to be used to select items of interest. Results will be presented in a future paper.

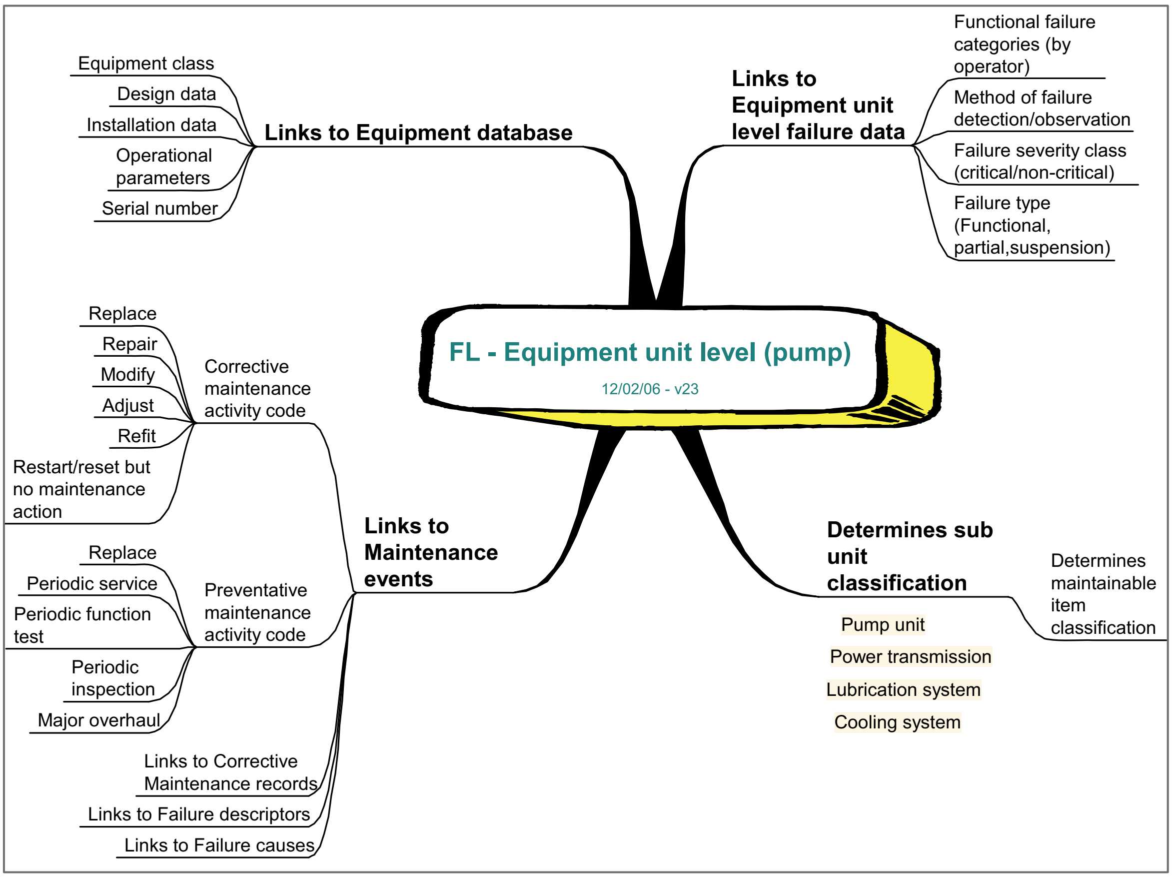


Figure 5: Illustration of the links between a functional location at the unit level and the fields in other parts of the asset management system.

Data ownership

In many ways it is unfortunate that the term "data ownership" was ever coined. The concept of data owner as a role implicitly assigns data maintenance tasks to the data owner. Therefore, by checking our own job descriptions to ensure that the words "data owner" are absent, we, as RAMS professionals, inevitably blame data quality problems on someone else. Yet, as the greatest users of RAMS data, our worth as analysts, engineers and professionals is directly linked to quality of the data on which we base our contribution. Ironically, the reticence of individuals to take on data ownership roles is often not due to an unwillingness to undertake the vetting incoming data, but rather a fear of opening a Pandora's Box of data quality issues that may then become a significant burden for an individual.

It is thus important to define responsibilities when assigning/accepting data ownership roles. Just because someone is prepared to maintain data does not necessarily mean they should be, individually, tasked with rectifying the sins of the past. We suggest that, where significant remediation or change to data management practice is identified, they are written up as separate business cases or project proposals. This helps establish the cost (and hopefully the value) of these changes in the context of the business needs.

5. CONCLUSIONS

There are a great many users of the same maintenance and failure data and considerable care should be taken before any changes are made to the data collection, data structure or data storage processes. This paper presents and illustrates a framework to assess the quality of data and the potential impact of data collection processes on the validity of key metrics used in assessing equipment reliability. We describe the new framework as a 'context-oriented' approach to data quality as the process assesses the data in individual fields in the context of its fitness for purpose; the purpose is the calculation of a specific reliability metric which must be necessary to support a specific business need.

Advantages of this approach include: (1) a structured prescriptive methodology that can be applied to many situations, (2) a direct link between data quality and the business process, and (3) identification of all data inputs and their associated data collector, data storage format and process. In the authors' opinion much more thought and attention needs to be given to the process of collecting, storing and using data for maintenance and reliability decisions. Using the framework described herein will assist in understanding assumptions used in reliability calculations and improve the quality of underlying data and data collection processes.

6. REFERENCES

[1] Wang, R. Y. & Strong, D. M. (1996) Beyond accuracy: What data quality means to consumers. *Journal of Management Information Systems*, 12(4), 5-34.

[2] Koronios, A. & Lin, S. (2004) Key issues in achieving data quality in Asset Management . VETOMAC-3/ACSIM-2004 (Vibration Engineering & Technology of Machinery, Asia-Pacific Conference on System Integrity & Maintenance 2004, December 6-9) New Delhi.

[3] Sandtorv, H. A., Hokstad, P. & Thompson, D. W. (1996) Practical experiences with a data collection project: the OREDA project. *Reliability Engineering and System Safety*, **51**, 159-167.

[4] Lee, Y.W. & Strong, D.M. (2004) Knowing-Why about Data processes and Data quality. *Journal of Management Information Systems*, 20(3), 13-39.

[5] International Standard: Petroleum and natural gas industries – Collection and exchange of reliability and maintenance data for equipment, ISO 14224 (1999).

[6] Hodkiewicz, M.R., Coetzee, J.L., Dwight, R.A. & Sharp, J.M. (2006) The importance of knowledge management to the Asset Management Process, *Business Briefings: Oil and Gas Processing* Review 2006, 43-45.

[7] Gilb, T. and Weinberg, G. (1977) Humanized Input: Technique for Reliable Keyed Input, *Winthrop Publishers, Inc*.

DETERMINATION OF OPTIMAL PREVENTIVE MAINTENANCE STRATEGY FOR SERIAL PRODUCTION LINES

Yong Sun[1], Lin Ma and Joseph Mathew

CRC for Integrated Engineering Asset Management, Shool of Engineering Systems, Queensland University of Technology, Brisbane, QLD 4001, Australia

Abstract: A Split System Approach (SSA) based methodology is presented to assist in making optimal Preventive Maintenance decisions for serial production lines. The methodology treats a production line as a complex series system with multiple PM actions over multiple intervals. Both risk related cost and maintenance related cost are factored into the methodology as either deterministic or random variables. This SSA based methodology enables Asset Management (AM) decisions to be optimized considering a variety of factors including failure probability, failure cost, maintenance cost, PM performance, and the type of PM strategy. The application of this new methodology and an evaluation of the effects of these factors on PM decisions are demonstrated using an example. The results of this work show that the performance of a PM strategy can be measured by its Total Expected Cost Index (TECI). The optimal PM interval is dependent on TECI, PM performance and types of PM strategies. These factors are interrelated. Generally it was found that a trade-off between reliability and the number of PM actions needs to be made so that one can minimize Total Expected Cost (TEC) for asset maintenance

Key Words: Preventive maintenance, Planning, Decision making, Production lines, Split System Approach

1 INTRODUCTION

The determination of optimal Preventive Maintenance (PM) strategies for production lines, especially over the whole life of these assets, is imperative for their owners as maintenance costs can occupy a sizeable portion of the total costs of business. The need to optimise maintenance of production lines becomes pressing with increasing complexity of machines and competitive market pressure.

Maintenance issues of production lines have attracted much attention from researchers. For example, Dallery and Bihan [1] developed an improved method for analysing serial production lines with unreliable machine and finite buffers. Liberopoulos [2] conducted a case study for the reliability analysis of an automated pizza production line and Miltenburg [3] investigated the effect of breakdown on U-shaped production lines. Some literature on the optimal PM planning for production lines has also been published. For example, see research reports presented by Cavory [4], Percy et al [5], and Chareonsuk et al [6]. Two major issues need to be addressed when making an optimal decision of PM strategy for production lines: (1) the changes in reliability of production lines due to PM and (2) maintenance related costs. Conflicting interests exist between these two issues. More frequent maintenance activities often need to be conducted and more resources need to be consumed if one wishes to maintain a production line at a higher reliability level. As a result, maintenance related costs increase. On the other hand, lowering reliability requirements can reduce the maintenance related costs. However, a lower reliability of a production line usually means that this production line is prone to more breakdowns and greater loss in production. A good maintenance strategy must balance both reliability and maintenance costs. Some analysis has revealed that maintenance cost will increase with increasing maintenance frequency whereas the cost due to breakdown of a production line decreases with increasing maintenance frequency. Hence, an optimal maintenance frequency exists [7]. However, this analysis was conducted qualitatively rather than quantitatively. For quantitative analysis, Chareonsuk et al [6] attempted to optimise PM intervals of production lines under two criteria – expected total costs per unit time and reliability. However, they did not consider multiple imperfect PM actions in their model. To deal with a long term PM schedule for new production lines, Percy et al [5] postulated a new Bayesian method based approach but did not develop an applicable algorithm. **A practical model for determining the optimal PM strategy for production lines over its life-span is yet to be developed**. The major barrier to develop such a model is reliability prediction of production lines with multiple PM actions over a long operational period. Production lines are

[1] Corresponding author: y3.sun@qut.edu.au, Tel: (61 7) 3864 2442, Fax: (61 7) 3864 1469

This research was conducted within the CRC for Integrated Engineering Asset Management, established and supported under the Australian Government's Cooperative Research Centres Programme.

normally complex repairable systems and PM actions on these complex systems are generally imperfect, i.e., the state of a production line after a PM action is between “as good as new” and “as bad as old”.

A Split System Approach (SSA) based methodology is developed in this paper to remove this barrier. SSA was proposed by the authors [8] recently to predict the reliability of systems with multiple PM actions over multiple intervals. In this paper, the SSA is used to predict the reliability of production lines with multiple PM actions. Only serial production lines are considered. A serial production line indicates that the failure of any machine in this production line will cause the failure of the whole system (production line). Serial production lines are commonplace in manufacturing industries such as automobile manufacturing factories, food processing factories and clothes making factories.

The rest of the paper is organised as follows: in Section 2, the concept and methodology of SSA are reviewed. In Section 3, a methodology for determining the optimal PM strategy based on SSA is presented, and this is followed by an example in Section 4. Conclusions are given in Section 5.

2 THE CONCEPT AND METHODOLOGY OF SSA

The basic concept of the SSA is to separate repaired and unrepaired components within a system virtually when modelling the reliability of a system after PM actions. This concept enables the analysis of system reliability at the component level, and stems from the fact that generally when a complex system has a PM action, only some of components are repaired.

The following assumptions were made in developing SSA based models:

The failure of repaired components is independent of unrepaired components. This assumption means that when a component is repaired, the failure distribution form of the unrepaired components of a system does not change, and the conditions of the unrepaired components do not affect the reliability characteristics of repaired components.

The reliability function of a new repairable system is known. The reliability functions of repaired components are also known.

The topology of a repairable system is known.

The repair time is negligible.

PM time is deterministic variable.

The topology of production lines discussed in this paper is assumed to be a serial system consisting of M components. The original multi-serial system can be converted into a simplified serial system which only contains two virtual parts: “Part 1” includes repaired machines and “Part 2” is the remainder of the production line - often referred to as a subsystem (see Figure 1).

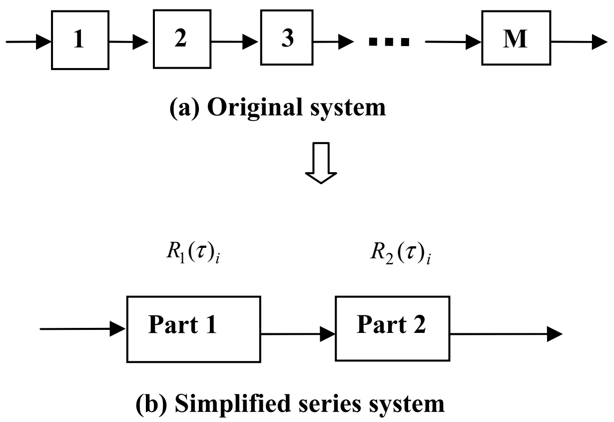


Figure 1. Simplification of production lines

In Figure 1, $R_1(\tau)_i$ and $R_2(\tau)_i$ are the reliability functions of Part 1 and Part 2 after the i[th] PM interval (refer to Figure 2). In this paper, the second subscript i is used to stand for “after the i[th] PM action”. Subscript $i=0$ stands for no PM. The PM strategy is to repair Part 1 whenever the reliability of the production line falls to a predefined control limit of reliability R_0. A possible interpretation for this PM strategy is that the components in Part 1 have a much shorter mean time to failure than the components in Part 2.

As mentioned previously, production lines are often complex repairable systems. The states of machines after repairs in a production line can have a significant impact on the reliability of the entire production line and must be considered while modelling the reliability of the production line covering a series of PM actions. PM actions on a production line often involve imperfect repairs. The reliability of a system after imperfect repairs declines in a manner shown in Figure 2.

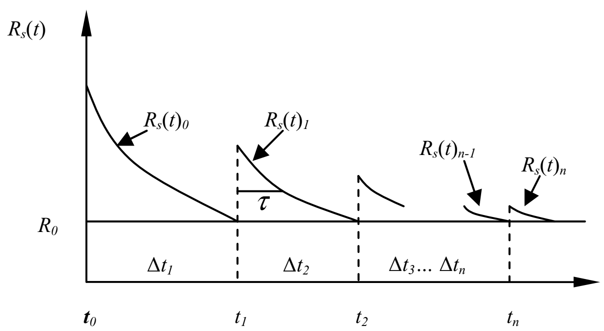


Figure 2. Changes to the reliability of an imperfectly repaired system

Two time coordinates are used in the modelling:

Absolute time scale t: $0 \le t < \infty$.

Relative time scale τ: $0 \le \tau \le t_i$ ($i = 1, 2, \ldots, n$).

In Figure 2, R_0 is the predefined control limit of the reliability level for the production line, Δt_i (i=1,2,…,n) is the interval time between (i-1)th repair and ith repair. Parameter t_i is the ith PM time and also the start time for the production to run again after the ith repair according to Assumption (4).

For a simple scenario that always Part 1 is repaired in all PM actions, the reliability function of the system after the nth PM actions can be expressed as:

$$R_s(\tau)_n = \frac{R_1(\tau)_n R_s(\tau + \sum_{i=1}^{n} \Delta t_i)_0}{R_1(\tau + \sum_{i=1}^{n} \Delta t_i)_0}. \quad (1)$$

Equation (1) can be rewritten using absolute time scale as follows:

$$R_s(t) = \frac{R_1(t - \sum_{i=1}^{n} \Delta t_i)_n R_s(t)_0}{R_1(t)_0} \qquad (t \geq \sum_{i=1}^{n} \Delta t_i). \quad (2)$$

Note that equations (1) and (2) both describe the reliability of a system which has been preventively maintained for n times, i.e., these two equations both describe the conditional probability of survival of a system with n PM intervals. To predict the reliability of a system with multiple PM intervals, the cumulative effect of multiple PM actions needs to be considered, i.e., the probability of survival of the repaired components until their individual repair times should be considered [7].

The cumulative reliability function of a serial system after the first PM action is

$$R_{sc}(\tau)_1 = R_1(\Delta t_1)_0 R_s(\tau)_1 \quad (3)$$

where, $R_{sc}(\tau)_1$ is the cumulative reliability of the system after the first PM action. $R_1(\Delta t_1)_0$ is the probability of survival of Part 1 until time t_1.

Generally, the cumulative reliability of the system after the nth PM action can be expressed as:

$$R_{sc}(\tau)_n = \prod_{i=0}^{n-1} R_1(\Delta t_{i+1})_i R_s(\tau)_n \quad (4)$$

where $R_{sc}(\tau)_n$ is the cumulative reliability of the system after the nth PM action.

3 Methodology for determining an optimal PM strategy

The SSA based PM decision making methodology is composed of reliability prediction of systems and cost analysis.

Estimation of the reliability of production lines

As mentioned in Section 2, SSA analyses the reliability of repairable systems after PM at the component level. Hence, direct application of SSA to estimating the reliability of production lines might be inconvenient because a production line often consists of numerous components. To avoid this inconvenience, a production line can be decomposed at different levels virtually, and then the reliability of the production line can be analysed at these levels using SSA respectively (see Figure 3).

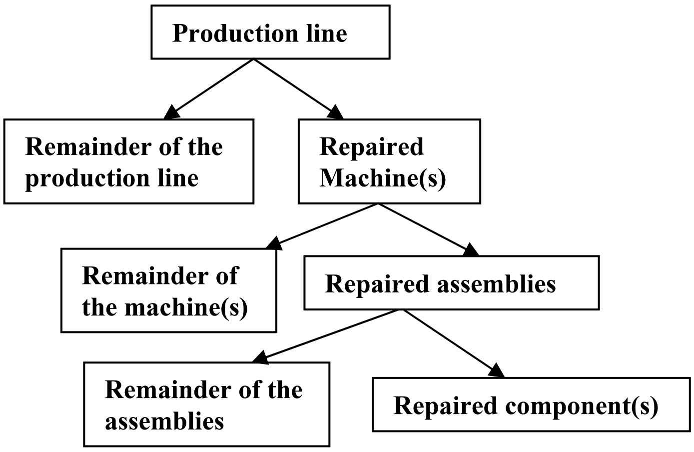


Figure 3. Decomposition of a production line

A bottom-up approach can be used for analysing the reliability of the production line after a production line has been virtually decomposed as shown in Figure 3. The reliability functions of assemblies are estimated firstly at the component level using SSA, and then the reliability functions of machines can be estimated at the assembly level. Finally the reliability function of the production line can be estimated at the machine level. For simplification, only the last step is demonstrated in this paper.

Criteria for optimizing PM strategies

Both reliability of a production line and maintenance related cost are considered in this paper when determining optimal PM strategies for production lines.

Reliability is used to describe the likelihood of failure of a system. Risk due to failure of production lines can be converted into risk related cost, which includes loss of production, penalty for contract breach, machine damages and additional harmful impact on human, products, machines and environment. Maintenance related cost includes material cost, maintenance labour cost and loss of production due to conducting PM. Various asset maintenance cost models have been developed (e.g., see references [9-11]). In this paper, the risk related cost of a production line is assumed to be proportional to the failure probability of the production line and the maintenance related cost is assumed to be proportional to the number of PM actions. Based on these two assumptions, the risk related cost and the maintenance related cost are expressed as:

$$C_r = k_r[1 - R(T)], \quad (5)$$

$$C_m = k_m N_T, \qquad (6)$$

where, T is the operational period of the production line that an enterprise is interested in. Typically, T is the life span of the production line. $R(T)$ is the reliability of the production line at time T. Parameters k_r and k_m are two scale constants. N_T is the required number of PM actions over the period of time T for maintaining the production line above the reliability level of $R(T)$.

Define the Total Expected Cost (TEC) as the sum of the expected risk related cost and the expected maintenance related cost and the Total Expected Cost Index (TECI) as the result that the TEC is divided by k_m:

$$TEC = C_r + C_m, \quad (7)$$

$$TECI = k_{rm}[1 - R(T)] + N_T, \quad (8)$$

where, $k_{rm} = k_r / k_m$. (9)

Parameter k_{rm} is termed as the Risk-Maintenance Cost Ratio (RMCR). It represents the significance of a PM action. A higher k_{rm} indicates that a PM action is more significant, that is, more risk related costs can be reduced due to the decreased failure probability after this PM action. An advantage for using RMCR is that this parameter is dimensionless.

TECI can be used to measure the performance of a PM strategy. The lower the TECI is, the better the PM strategy.

In industry, parameters k_r and k_m may vary significantly and unpredictably. Let K_r denote the cost per PM action and K_m denote the cost per percentage of failure probability. Then K_r and K_m are both random variables. Assume that K_r and K_m both change in $[0, \infty)$ and are independent of the age of the asset and the number of PM actions. If K_r has a distribution density function $f_r(k_r)$, then conditional on $K_r = k_r$, one has

$$C_r[R(T) \mid K_r = k_r] = k_r[1 - R(T)] \quad (10)$$

and on removing the condition, one has

$$C_r = \int_0^\infty k_r[1 - R(T)] f_r(k_r) dk_r$$

$$= E[K_r][1 - R(T)] \text{ ,(11)}$$

where, $E[K_r] = \int_0^\infty k_r f_r(k_r) dk_r$ is the first moment of K_r.

Similarly, if K_m has a distribution density function $f_m(k_m)$, the expected maintenance cost is given by

$$C_m = E[K_m] N_T \text{ , } \quad (12)$$

where, $E[K_m] = \int_0^\infty k_m f_m(k_m) dk_m$ is the first moment of K_m.

In this case, define RMCR as

$$k_{rm} = \frac{E[K_r]}{E[K_m]}, \qquad (13)$$

so that equation (8) still holds.

The approach to determining the optimal PM strategy for production lines presented in this section is best demonstrated using an example in the following section.

4 EXAMPLE

A PM strategy is required for a period of the next two years for an automated food production line that has been operating for some time. This production line can be described as a simplified serial system as shown in Figure 1. Part 1 is composed of those machines that have very short mean time to failure compared with the remainder of the production line and Part 2 is composed of the remainder of the production line. This production line was assumed to operate at its wear-out stage during the next two years. The times of critical failures of Part 1 followed a Weibull distribution and were expressed as:

$$R_1(\tau)_0 = \exp[-\left(\frac{\tau}{18}\right)^{2.1}]. \qquad (14)$$

Part 2 was assumed to have an exponential failure distribution:

$$R_2(\tau)_0 = \exp\left(\frac{-\tau}{400}\right). \qquad (15)$$

Hence the reliability of the entire production line was

$$R_s(\tau)_0 = \exp[-\left(\left(\frac{\tau}{18}\right)^{2.1} + \left(\frac{\tau}{400}\right)\right). \quad (16)$$

Conducting PM on the machines in Part 1 can improve the overall reliability of the entire production line since Part 1 was operating at its wear-out stage. This scenario has been studied in subsection 3.2. The reliability of the entire production line with multiple PM intervals can be analysed using Equations (1) and (4).

Two PM strategies were considered. Strategy one is a type of Reliability Based PM (RBPM) strategy. In this strategy, Part 1 will be maintained whenever the reliability of the entire production line after PM falls to 0.9. The required minimum operational time of the production line after a PM action is 0.5 months (15 days. A calendar system of twelve 30-day months is used in this paper.) The second strategy is a type of Time Based PM (TBPM) strategy. In this PM strategy, PM on the machines in Part 1 starts after one month (30 days) after which it will be conducted in fixed intervals. As mentioned in Section 3.1, the reliability of Part 1 after maintenance can also be predicted using the SSA. However, the derived reliability formula is complicated. In this paper, the following approximate formula was used to describe the reliability of Part 1 after a repair:

$$R_1(\tau)_n = R_1(\tau + f_c \Delta t_{n-1}), \quad (17)$$

where, f_c is termed as the recovery coefficient, which is used to represent the degree of the reliability of Part 1 after a PM action has recovered to its original reliability. When $f_c = 0$, the state of Part 1 after a PM action is as good as new; When $f_c = 1$, the state of Part 1 after a PM action is as bad as old; When $0 < f_c < 1$, Part 1 has an imperfect repair.

The reliability of the entire production line with different PM strategies was predicted using SSA. Two examples of reliability prediction are shown in Figures 4 and 5. In both figures, $f_c = 0.05$. From these two figures, it can be seen that both TBPM and RBPM strategies improved the reliability of the entire production line. One can find that the reliability of the production line with a shorter PM interval is higher, when considering the cumulative reliability with TBPM solely. However, this result does not mean that PM strategy with a shorter interval is superior over the PM strategy with a longer interval because the required number of PM actions corresponding to the shorter PM interval is higher than that corresponding to the longer PM interval during the same period. More PM actions often cause higher maintenance costs. An optimal decision of PM strategy should be based on both reliability requirement and maintenance costs. A trade-off between reliability level and the number of PM actions is necessary to keep the TEC at the lowest level.

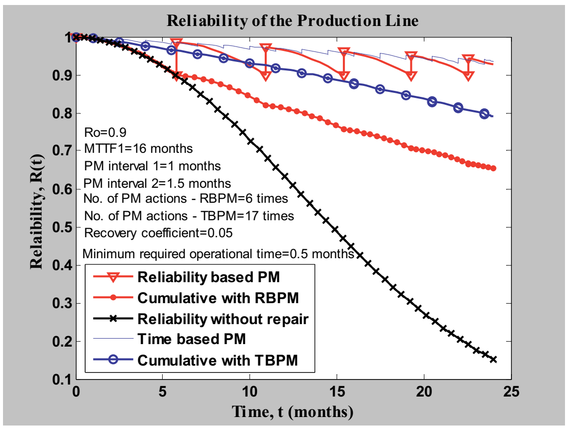


Figure 4. Reliability prediction of the production line – simulation 1

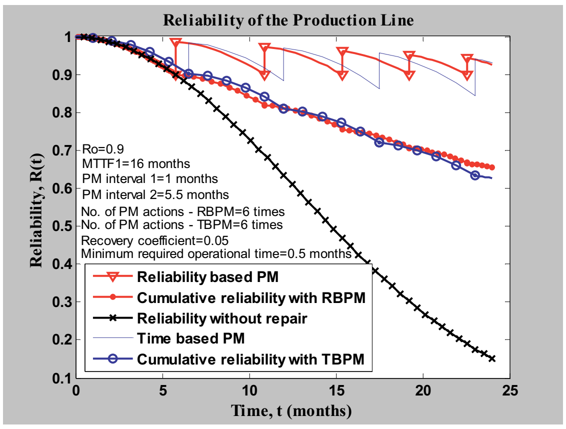


Figure 5. Reliability prediction of the production line – simulation 2

An optimal PM interval exists (see Figure 6). This optimal PM interval is RMCR, k_{rm}, dependent. From Figure 6, it can be seen that the best PM interval is two months when RMCR k_{rm} are 200 and 100. When k_{rm} is 10, the optimal PM interval changes to 8 months. However, when RMCR k_{rm} is 4, the optimal PM interval becomes 24 months. This result indicates that PM is no longer needed in this case because the risk related cost is not significant compared with the maintenance related cost.

The optimal interval is also dependent on the recovery coefficient f_c. From Figure 7, it can be seen that the optimal interval increases with the increase of recovery coefficient. When f_c is greater than 0.75, the optimal PM interval becomes 24 months, that is, no TBPM is required during the scheduled operating period of the production line. This finding can be explained by the property of the recovery coefficient. As shown in Equation (12), the recovery coefficient f_c represents the degree that the reliability of Part 1 after a PM action lowers its initial reliability before this PM. In other words, the recovery coefficient f_c represents the effectiveness of a PM action. A greater value of f_c indicates poorer PM performance. If PM performance is so degraded that it cannot improve the reliability of production lines effectively, it is better that this PM not be conducted.

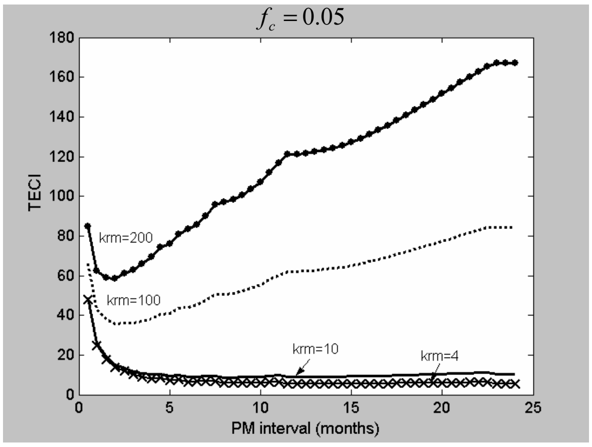


Figure 6. Relationship between TECI and preventive maintenance intervals

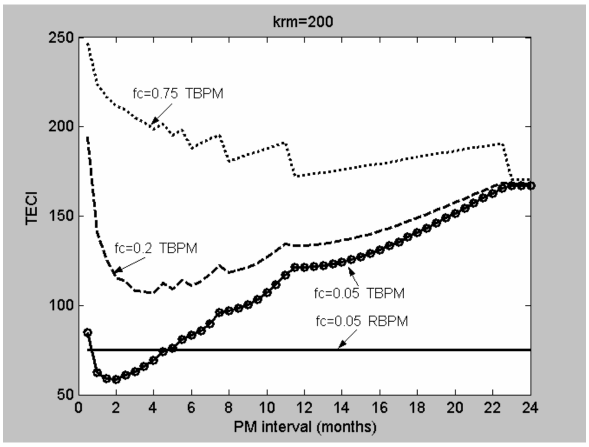


Figure 7. Relationship between TECI and the recovery coefficients

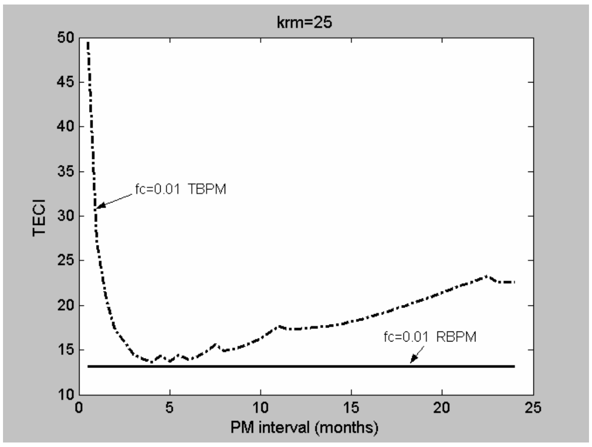


Figure 8. Comparison between RBPM and TBPM

f_c	Optimal PM interval (months)	Lowest TECI		k_{rm}
		TBPM	RBPM	
0.05	2	58.6	75.1	200
0.1	2	75.9	85.5	
0.15	3	92.2	108	
0.2	4	106.9	inapplicable	
0.3	6	129.9	inapplicable	
0.7	11.5	169.4	inapplicable	

Table 1. Relationship between TECI and the recovery coefficients

The above analysis focuses on obtaining optimal TBPM strategy. However, there are times when another type of PM strategy can be better than this optimal TBPM. When determining an optimal PM strategy, one needs to investigate different types of PM strategies. The effectiveness of different types of PM strategies can vary in different scenarios. In the scenario shown in Figure 8, the lowest TECI for TBPM is 13.7 whereas the TECI for RBPM is 13.1, i.e., in this case, RBPM strategy rather than TBPM strategy should be applied. However, in the scenario presented in Table 1, TBPM strategy is better than RBPM strategy.

In Table 1, the word “inapplicable” means that RBPM is not applicable because the PM interval required by this strategy will become shorter than the required minimum operational time of the production line.

5 CONCLUSIONS

A SSA based methodology for determining an optimal PM strategy of production lines was developed in this paper. This methodology is especially useful for long term PM decision making.

The determination of an optimal PM strategy of production lines is essentially a multiple criteria decision making issue. A number of factors can influence production line PM decision making. The major factors include failure probability, costs due to failure of production lines, costs relating to maintenance, PM performance, and the type of PM strategy. The SSA based methodology considers all these factors simultaneously and analyses the effects of these factors on PM decisions quantitatively.

This research work finds that the performance of a PM strategy can be measured by its Total Expected Cost Index (TECI). The lower the TECI, the better the PM strategy. The effectiveness of different types of PM strategies can vary in different scenarios. The optimal PM interval is dependent on TECI, PM performance and the type of PM strategy. A trade-off between reliability requirement and the number of PM actions is often needed if one wishes to minimise the Total Expected Cost (TEC) of using production lines.

6 REFERENCES

1 Dallery Y and Bihan HL. (1999) An improved decomposition method for the analysis of production lines with unreliable machines and finite buffers. Int. J. of Production Research, *37*(5), 1093-1117.

2 Liberopoulos G and Tsarouhas P. (2004) Reliability analysis of an automated pizza prodiction line. J. of Food Engineering, In press.

3 Miltenburg J. (2000) The effect of breakdowns on U-shaped production lines. Int. J. of Production Research, *38*(2), 352-364.

4 Cavory G, Dupas R and Goncalves G. (2001) A genetic approach to the scheduling of preventive maintenance tasks on a single product manufacturing production line. int. J. of Production Economics, *74*(1), 135-146.

5 Percy DF, Kobbacy KAH and Fawzi BB. (1997) Setting preventive maintenance schedules when data are sparese. int. J. of Production Economics, *51*(2), 223-234.

6 Charepnsuk C, Nagarur N and Tabucanon MT. (1997) A multicriteria approach to the selection of preventive maintenance intervals. int. J. of Production Economics, *49*(1), 55-64.

7 Ebeling CE. (1997) An Introduction to Reliability and Maintainability Engineering. New York: The McGraw-Hill Company, Inc.

8 Sun Y, Ma L and Mathew J. (2004) Reliability prediction of repairable systems for single component repair. In J. Lee. (Eds) *Proceedings of International Conference on Intelligent Maintenance System,* Arles, France. pp. S2-A. IMS.

9 Kelly A. (1984) Maintenance Planning and Control. Cambridge: Butterworth & Co Ltd.

10 Pham H (ed.) (2003), Handbook of Reliability Engineering, London: Springer

11 Blischke WR and Murthy DNP. (2000) Reliability - Modelling, Prediction, and Optimization. New York: John Wiley &Sons, Inc.

THE INTERACTIVE HAZARDS OF COMPONENTS IN A SYSTEM AFTER REPAIRS – A MATHEMATICAL FORMULATION AUTHORS

Y Sun L Ma & J Mathew

CRC for Integrated Engineering Asset Management,School of Engineering Systems,Queensland University of Technology,Brisbane, QLD 4001, Australia

Abstract: This paper develops analytical expressions that enable analysis of the changes of component hazards of systems with interactive failures after repairs. These changes were analysed using probability theory based on the analytical model for interactive failures. It was found that hazards for repaired and *unrepaired* components will change when a system with interactive failures is maintained. The interactive hazards of newly repaired components are higher than their independent hazards; whereas the interactive hazards of unrepaired components will reduce. When two components interact strongly, these two components should be maintained at a similar condition. When establishing effective asset maintenance management strategies one needs to understand the changes in hazards of components when a system with interactive failures is maintained. The proposed model calculates and predicts these changes in the hazard. This work is the first application of the analytical model for interactive failures in the modelling of changes in component hazards of a system after repairs. The outputs of this work can be used to optimise maintenance strategies for systems with interactive failures.

Key Words: hazard, interactive failure, repairable systems, asset maintenance.

1 INTRODUCTION

The majority of physical assets in industry such as machines, buildings and vehicles are repairable complex systems. These types of assets will need to be repaired to recover their functions after each failure. Optimal maintenance of repairable assets is important to enterprises for delivering efficiencies. The development of an optimal maintenance strategy for these assets requires the accurate prediction of their reliability after repairs.

Numerous models and methodologies have been developed to meet this requirement. However, most these models and methodologies are based on the assumption of independent failures. Industrial experience has shown that there are a number of situations where this assumption is unrealistic and will lead to unacceptable analysis errors. To ensure the accuracy of reliability prediction in these situations, dependent failures must be considered. Interactive failure is a type of dependent failure and is typically caused by interactions between components in a system.

The authors' previous research [1] revealed that interactions between components will increase the hazards of the components. This increased hazard is defined as Interactive Hazard (IntH). Correspondently, the hazard of a system without failure interactions is termed as Independent Hazard (IndH). Several researchers [2] have previously observed that the interactive hazards of both repaired and unrepaired components in a system will change after repairs. Understanding these changes of hazards can assist in developing an optimal maintenance strategy for the system. However, an analytical analysis of these changes is yet to be conducted.

This paper quantitatively analyzes the changes of component interactive hazards after repairs using the Analytical Model for Interactive Failures (AMIF) previously developed by the authors [1].

The remainder of this paper is organized in the following manner: In Section 2, the changes of component hazards are described. A quantitative analysis of these changes is presented in Section 3. Section 4 presents the conclusions.

2 CHANGES OF COMPONENT HAZARDS

The reliability of a system is expected to increase after repair because the hazard of this system is reduced. In case of a system with independent failures, only the hazards of repaired components are assumed to be changeable, and the hazards of unrepaired components after a repair are assumed unchangeable. However, if a system has interactive failures, the hazards of both repaired and unrepaired components are changeable. In this case, repairs can improve the reliability of a system in two ways: reducing the IntH of unrepaired components and reducing the IndH of repaired components (refer to Figure 1).

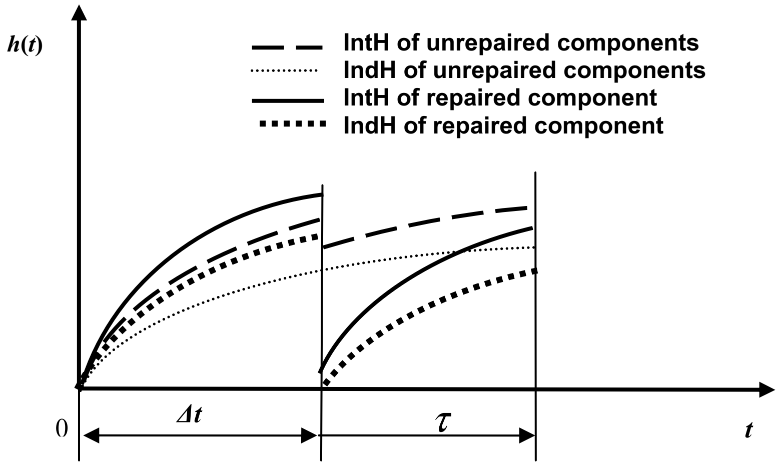


Figure 1. The changes of hazard of unrepaired components and repaired component

In Figure 1, parameter t represents the absolute time scale which changes from zero to infinite, whereas τ represents the relative time scale, which changes from zero to Δt. Parameter Δt is the interval between two repairs. The initial time to calculate the IndH of the unrepaired components after a repair is t_1 and the initial time to calculate the IndH of the newly repaired component after this repair is zero.

The changes of component hazards of a system after repairs is analysed below.

3 MODEL FORMULATION

This paper focuses on calculating the hazards of the components in a system with IntF after a repair. The following assumptions are made:

1. The system is composed of M components and one specific component (Component 1) is repaired.
2. The interactive coefficients are constant and independent of repairs.
3. Repair time is neglected.

In the case of repairable systems with IntF, the initial time for calculating the IntH of newly repaired components can be different from that for remaining unrepaired components after a repair (see Figure 1).

According to [3], the stable IntH of a system is given by

$$\{h(\tau)\} = [\alpha]\{h_I(\tau)\}. \quad (1)$$

where, $\{h(\tau)\}$ is the stable IntHs of a system after failure interaction. It is an $M \times 1$ vector. $\{h_I(\tau)\}$ is an $M \times 1$ independent hazard vector of all components due to their own deteriorations. $[\alpha]$ is the State Influence Matrix (SIM) which is given by

$$[\alpha] = ([I]-[\theta])^{-1}. \quad (2)$$

where, $[I]$ is an $M \times M$ unit matrix. $[\theta]$ is the Interactive Coefficient (IC) matrix of the system [1].

Define all unrepaired components as a subsystem. Equation (1) can be rewritten using the partition matrix as follows:

$$\begin{Bmatrix} h_1(\tau) \\ \vec{h}_{sb}(\tau) \end{Bmatrix} = \begin{bmatrix} \alpha_{11} & \vec{\alpha}_2 \\ \vec{\alpha}_3 & \vec{\alpha}_4 \end{bmatrix} \begin{Bmatrix} h_{I1}(\tau) \\ \vec{h}_{Isb}(\tau) \end{Bmatrix}. \quad (3)$$

where, $h_1(\tau)$ is the IntH of Component 1. Vector $\vec{h}_{sb}(\tau)$ is the $(M-1)\times 1$ IntH vector of the subsystem. Parameter α_{11} is the first row first column element of SIM $[\alpha]$; while $\vec{\alpha}_2$, $\vec{\alpha}_3$ and $\vec{\alpha}_4$ are the $1\times(M-1)$, $(M-1)\times 1$ and $(M-1)\times(M-1)$ partition matrix in SIM $[\alpha]$, respectively. Function $h_{I1}(\tau)$ is the IndH of Component 1, and $\vec{h}_{Isb}(\tau)$ is a $(M-1)\times 1$ vector which represents the IndH of the subsystem.

Let $h_{I1}(\tau)_0$ and $\vec{h}_{Isb}(\tau)_0$ denote the IndH of Component 1 and the subsystem before a repair respectively.

When a repair is conducted, $\tau = t_1 = \Delta t$. Hence, just before this repair, the IndHs of Component 1 and the subsystem are $h_{I1}(\Delta t)_0$ and $h_{Isb}(\Delta t)_0$, respectively. Let $h_{I1}(\tau)_1$ be the IndH of Component 1 after this repair, and then the IndH of Component 1 just after the repair is $h_{I1}(0)_1$. Generally

$$0 \le h_{I1}(0)_1 \le h_{I1}(\Delta t_1)_0 . \qquad (4)$$

If $h_{I1}(\tau)_1 = h_{I1}(\tau + \Delta t_1)_0$, the state of the system after the repair is "as bad as old".

The IndH of the subsystem just after a repair is the same as just before this repair because it has not been repaired, i.e.,

$$h_{Isb}(\tau)_1 = h_{Isb}(\tau + \Delta t_1)_0 , \qquad (5)$$

where, $h_{Isb}(\tau)_0$ and $h_{Isb}(\tau)_1$ are the IndHs of the subsystem before and after the repair respectively.

The IntHs of all components in the system after the repair are given by

$$\begin{Bmatrix} h_1(\tau)_1 \\ \vec{h}_{sb}(\tau)_1 \end{Bmatrix} = \begin{bmatrix} \alpha_{11} & \vec{\alpha}_2 \\ \vec{\alpha}_3 & \vec{\alpha}_4 \end{bmatrix} \begin{Bmatrix} h_{I1}(\tau)_1 \\ \vec{h}_{Isb}(\tau + \Delta t_1)_0 \end{Bmatrix}, \quad (6)$$

where, $h_{I1}(\tau)_1$ is the IndH of Component 1 after a repair. Vector $\vec{h}_{sb}(\tau)_0$ represents the IntHs of Component 1 and the subsystem before the repair; while $h_1(\tau)_1$ and $\vec{h}_{sb}(\tau)_1$ are the IntHs of Component 1 and the subsystem after the repair respectively.

If IntF is stable and the condition of Component 1 just after the repair has not degraded compared to its condition just before this repair, the following inequalities can be obtained:

$$h_1(t_1)_0 \ge h_1(0)_1 = \alpha_{11} h_{I1}(0)_1 + \vec{\alpha}_2 \vec{h}_{Isb}(\Delta t_1)_0 \ge h_{I1}(0)_1 , \qquad (7)$$

$$\vec{h}_{sb}(0)_1 = \vec{\alpha}_3 h_{I1}(0)_1 + \vec{\alpha}_4 \vec{h}_{Isb}(\Delta t_1)_0 \le \vec{h}_{Isub}(\Delta t_1)_0 . \quad (8)$$

The above inequalities can be proved using the following two propositions and a theorem.

Proposition 1: All elements in SIM $[\alpha]$ are nonnegative when $0 \le \theta_{ij} < 1$.

The proof of Proposition 1 is presented in Appendix 1.

Proposition 2: All diagonal elements in SIM $[\alpha]$ are greater than or equal to one.

The proof of Proposition 2 is presented in Appendix 2.

Theorem: Interactive functions $h_1(\tau)$ and $\vec{h}_{sb}(\tau)$ change monotonously with the change of $h_{I1}(\tau)$.

The proof of this theorem is straightforward using Equation (3) and Proposition 1.

Eq (7) is proved as follows:

According to Proposition 1, $\vec{\alpha}_2 \ge 0$. According to Proposition 2, $\alpha_{11} \ge 1$. Hence, the following relationship holds because all elements in $\vec{h}_{Isb}(\Delta t_1)_0$ are nonnegative:

$$h_1(0)_1 = \alpha_{11} h_{I1}(0)_1 + \vec{\alpha}_2 \vec{h}_{Isb}(\Delta t_1)_0 \ge h_{I1}(0)_1 . \qquad (9)$$

If the condition of Component 1 just after the repair has not degraded compared to its condition just before this repair, i.e., $h_{I1}(0)_1 \le h_{I1}(\Delta t_1)_0$, the following inequality holds because of Equation (6) and the theorem:

$$h_1(t_1)_0 \ge h_1(0)_1 . \qquad (10)$$

Equation (7) is obtained by a combination of Equations (9) and (10). Equation (8) can be proved using a similar approach.

Equation (9) indicates that the IntH of Component 1 can be higher than its original independent hazard due to the effect of the unrepaired subsystem. The inequity symbol in Equation (9) becomes the equality symbol if and only if $\vec{\alpha}_2$ is a null vector. A null vector $\vec{\alpha}_2$ means that the failures of components in a subsystem do not affect the failure of Component 1. If $\vec{\alpha}_2$ is a null vector, element α_{11} is equal to one (see Appendix 1). Equation (8) indicates that the IntHs of the components in the subsystem, and hence the subsystem, will be reduced after the repair. The inequality symbol in Equation (8) becomes the equality symbol if and only if $\vec{\alpha}_3$ is a null vector. A null vector $\vec{\alpha}_3$ would mean that the failure of Component 1 does not influence the failure of components in the subsystem.

4 CONCLUSIONS

This study has found that the initial time used in the calculation of the interactive hazards of repaired components after a repair is different from that of the remaining unrepaired components after this repair for a system with interactive failures. Repair can improve the reliability of a system in two aspects: decreasing interactive hazards of the unrepaired components and increasing the reliability of repaired components. The interactive hazards of newly repaired components are higher than their independent hazards; whereas the interactive hazards of unrepaired components will reduce. When two components interact strongly, these two components should be maintained at a similar level for effective maintenance. This paper models the changes in hazards of components in a system after a single repair. The authors are currently expanding on this work to cover systems with multiple repairs.

5 REFERENCES

[1] Sun Y, Ma L, Mathew J and Zhang S. (2006) An analytical model for interactive failures. Reliability Engineering & System Safety, 91(3), 495-504.

[2] Wang KS, Po, H.J., Hsu, F.S., and Liu, C.S. (2002) Analysis of equivalent dynamic reliability with repairs under partial information. Reliability Engineering & System Safety, 76(1), 29-42.

[3] Sun Y. (to be published). Reliability Prediction of Complex Repairable Systems. Queensland University of Technology

[4] Courant R and Bobbins H. (1996) What is Mathematics? New York: Oxford University Press.

6 ACKNOWLEDGEMENTS

Dr. Myung-Hwan Na from the Department of Statistics at Chonnam National University assisted in checking the mathematical models.

Appendix 1

The Proof of Proposition 1

Proposition 1: All elements in the State Influence Matrix (SIM) $[\alpha]$ are nonnegative when $0 \le \theta_{ij} < 1$.

Proof

Proposition 1 is proven using the Principle of Mathematical Induction [4] as follows.

According to reference [1], SIM $[\alpha]$ is the inverse matrix of $([I]-[\theta])$:

$$[\alpha] = ([I]-[\theta])^{-1}, \qquad \text{(A1-1)}$$

where,

$$([I]-[\theta]) = \begin{bmatrix} 1 & -\theta_{12} & \cdots & -\theta_{1M} \\ -\theta_{21} & 1 & \cdots & -\theta_{2M} \\ \vdots & \cdots & \ddots & \vdots \\ -\theta_{M1} & -\theta_{M2} & \cdots & 1 \end{bmatrix}. \qquad \text{(A1-2)}$$

M is the number of components in a system. Matrix (A1-2) has the following properties:

1. All diagonal elements are equal to 1.
2. All non-diagonal elements are either negative or zero because

$$0 \le \theta_{ij} < 1 \quad (i, j = 1,2,...,M;\ i \ne j). \qquad \text{(A1-3)}$$

When $M = 2$,

$$[\alpha] = ([I]-[\theta])^{-1} = \begin{bmatrix} 1 & -\theta_{12} \\ -\theta_{21} & 1 \end{bmatrix}^{-1}$$

$$= \frac{1}{\det[([I]-[\theta])]} \begin{bmatrix} 1 & \theta_{12} \\ \theta_{21} & 1 \end{bmatrix}. \qquad \text{(A1-4)}$$

The proposition is true because $\det[([I]-[\theta])] > 0$.

Suppose that the proposition is true when $M = K$, i.e.,

$$\alpha_{ij} \ge 0 \quad (i, j = 1,2,\ldots,K). \qquad \text{(A1-5)}$$

When $M = K+1$, we can rewrite matrix $[\alpha]$ in the form of partition matrix:

$$[\alpha] = \begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1k} & \alpha_{1K+1} \\ \vdots & \ddots & \vdots & \vdots \\ \alpha_{K1} & \cdots & \alpha_{KK} & \alpha_{KK+1} \\ \alpha_{K+11} & \cdots & \alpha_{K+1K} & \alpha_{K+1K+1} \end{bmatrix} = \begin{bmatrix} \vec{\alpha}_{11} & \vec{\alpha}_{12} \\ \vec{\alpha}_{21} & \alpha_{K+1K+1} \end{bmatrix}. \qquad \text{(A1-6)}$$

In Equation (A1-6),

$$\vec{\alpha}_{11} = \begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1K} \\ \vdots & \ddots & \vdots \\ \alpha_{K1} & \cdots & \alpha_{KK} \end{bmatrix}, \qquad \text{(A1-7)}$$

$$\vec{\alpha}_{12} = \{\alpha_{1K+1}, \alpha_{2K+1}, \ldots, \alpha_{KK+1}\}^T, \qquad \text{(A1-8)}$$

$$\vec{\alpha}_{21} = \{\alpha_{K+11}, \alpha_{K+12}, \ldots, \alpha_{K+1K}\}. \qquad \text{(A1-9)}$$

Rewrite the matrix $([I]-[\theta])$ into the same sized partition matrix. Let $[\nu] = [I]-[\theta]$, then

$$[\nu] = \begin{bmatrix} 1 & -\theta_{12} & \cdots & -\theta_{1K+1} \\ -\theta_{21} & 1 & \cdots & -\theta_{2K+1} \\ \vdots & \cdots & \ddots & \vdots \\ -\theta_{K+11} & -\theta_{K+12} & \cdots & 1 \end{bmatrix} = \begin{bmatrix} \vec{\nu}_{11} & \vec{\nu}_{12} \\ \vec{\nu}_{21} & 1 \end{bmatrix}, \qquad \text{(A1-10)}$$

where,

$$\vec{v}_{11}=\begin{bmatrix} 1 & -\theta_{12} & \cdots & -\theta_{1K} \\ -\theta_{21} & 1 & \cdots & -\theta_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ -\theta_{K1} & -\theta_{K2} & \cdots & 1 \end{bmatrix}, \qquad \text{(A1-11)}$$

$$\vec{v}_{12}=\{-\theta_{1K+1},-\theta_{2K+1},\ldots,-\theta_{kK+1}\}^T, \qquad \text{(A1-12)}$$

$$\vec{v}_{21}=\{-\theta_{K+11},-\theta_{K+12},\ldots,-\theta_{K+1K}\}. \qquad \text{(A1-13)}$$

The following equation can be obtained by using the equation $[\alpha][\nu]=[I]$ and matrix multiplying rules:

$$\vec{\alpha}_{11}\vec{v}_{12}+\vec{\alpha}_{12}=\{0\}, \qquad \text{(A1-14)}$$

where, $\{0\}$ is a $1\times K$ null vector.

From Equation (A1-14), one can obtain the following equations:

$$-\sum_{s=1}^{K}\alpha_{is}\theta_{sK+1}+\alpha_{iK+1}=0 \quad (i=1,2,\ldots,K). \qquad \text{(A1-15)}$$

The first term in Equation (A1-15) is equal to or less than zero because of Equations (A1-3) and (A1-5). Therefore,

$$\alpha_{iK+1}\geq 0 \quad (i=1,2,\ldots,K). \qquad \text{(A1-16)}$$

On the other hand,

$$[\nu][\alpha]=[I]. \quad \text{(A1-17)}$$

Then the following result can be gained by using the same inference as mentioned above:

$$-\sum_{s=1}^{K}\theta_{K+1s}\alpha_{sj}+\alpha_{K+1j}=0 \quad (j=1,2,\ldots,K). \qquad \text{(A1-18)}$$

From Equation (A1-18), one has

$$\alpha_{K+1j}\geq 0 \quad (j=1,2,\ldots,K). \qquad \text{(A1-19)}$$

Furthermore, from

$$\alpha_{K+1K+1}-\sum_{s=1}^{K}\theta_{K+1s}\alpha_{sK+1}=1, \quad \text{(A1-20)}$$

the following conclusion can be drawn:

$$\alpha_{K+1K+1}\geq 1. \quad \text{(A1-21)}$$

A combination of Inequalities (A1-16), (A1-19) and (A1-21) gives

$$\alpha_{ij}\geq 0 \quad (i,j=1,2,\ldots,K+1). \quad \text{(A1-22)}$$

That is, when $M=K+1$, the Proposition 1 is also true.

Appendix 2

The Proof of Proposition 2

Proposition 2: All diagonal elements in the State Influence Matrix $[\alpha]$ are greater than or equal to one.

Proof

According to Equation (A1-17),

$$\alpha_{ii} - \sum_{\substack{s=1 \\ s\neq i}}^{M} \theta_{is}\alpha_{si} = 1. \qquad \text{(A2-1)}$$

The second term on the left side of Equation (A2-1) is not negative according to the properties of Interactive Coefficient (IC) and Proposition 1. Therefore,

$$\alpha_{ii} \geq 1 \quad (i = 1, 2, \cdots, M). \qquad \text{(A2-2)}$$

The inequality symbol becomes the equal symbol if all θ_{is}=0 ($s = 1, 2, \ldots, M$).

Propositions 1 and 2 have explicit physical meanings. Proposition 1 indicates that components in a system are subject to stable IntF. Proposition 2 indicates that the IntHs of the affected components in a system are greater than their Independent Hazards (IndHs) due to failure interactions. The failure likelihoods of these affected components also increase. The IntH of a component will be equal to its IndH if the failures of other components do not affect it.

BEHAVIOURAL PREFERENCES FOR ENGINEERING ASSET MANAGEMENT

J Amadi-Echendu

Dept of Engineering and Technology Management, University of Pretoria, Pretoria 0002, South Africs

Abstract: The essence of managing physical assets that form our built environment is to provide benefits to satisfy the continuum of constraints imposed by rapidly changing business strategy, economy, ergonomics, operational and technical integrity, and regulatory compliance. Innovative approaches to enhance and sustain the profile of values required from these assets demands a shift in the behavioural preferences or attitudes of engineering professionals in asset management occupations. This implies that engineering professionals in asset management occupations need to adapt to new thinking styles, and adopt effective cognitive and mental processing modes. Whilst assuming that thinking styles manifest in attitudes, this paper describes the results of a 2005 survey of 190 practicing engineers to ascertain what thinking styles should determine behavioural preferences of engineering-oriented managers of built environment assets of the innovation generation. The study confirms other results from cognitive theory and psychology, highlighting the top ten thinking styles as ranked by survey respondents. The paper is intended to provide a strategic view of Engineering Asset Management within the context of innovation, with particular focus on behavioural alignment towards the modern paradigm of the knowledge and learning economy.

Key Words: Asset Management, Thinking Styles, Behavioural Preferences, Innovation

1 INTRODUCTION

The main elements of our built environment are engineered physical assets which include, for example, airports, buildings, manufacturing and process plant, power stations, road and railway networks and systems, oil, gas and mining installations, telecommunications and utility networks. From an accounting and financial management point-of-view, these assets generally fall into four broad categories – (i) plant and equipment, (ii) buildings and infrastructure, (iii) furniture and fittings, and (iv) information technology. Overall management of built environment assets involves (i) acquisition, (ii) ownership, (iii) control, and (iv) utilisation. The essence of management is to ensure that the value profile defined by all stakeholders, is enhanced in a sustainable manner throughout the asset's life. The value that an asset can provide may be a broad combination of economic and social benefits depending on the preference and composition of the stakeholders. Acquisition, ownership, control and utilisation are high level management processes necessary to satisfy the continuum of constraints imposed by business strategy, economy, ergonomics, operational and technical integrity, and regulatory compliance [1].

From a systems viewpoint, the life of a built environment asset may be described interms of the phases and stages [2] illustrated in Figure 1. The two broad stages of capital development and business operations may be subdivided into four phases and further resolved into dominant activity periods. Successful management of the asset depends on innovative and synergistic integration of a wide range of 'hard' and 'soft' skills through the life-cycle phases and stages. These skills include, for example accounting, communications, engineering and technology disciplines. Whilst a broad range of skills is necessary, however, behavioural preferences or attitudes really underpin effective management of physical assets.

This paper examines the behavioural preferences of engineering professionals as physical asset managers within the context of innovation. In addition to managing the value profile defined by stakeholders at the various levels of the asset heirarchy, *Engineering Asset Managers* also have the responsibility to manage process innovation associated with, as well as technological innovation embedded in built environment assets.

2 THINKING STYLES AND BEHAVIOURAL PREFERENCES

With the advent of the new paradigm of innovation economy, knowledge has become the primary means of production. Thus, the knowledge base and skill set of engineering professionals are prime resources for effective management of physical assets in the new dispensation. *Engineering Asset Managers* form an important *ethnic* group of the innovation generation since they are vital actors in the processes of knowledge creation, diffusion, and transformation. Therefore, the behavioural preferences of engineering professionals affect the motivation, practice, roles and responsibilities, as well as the effectiveness of physical asset management.

From the point-of-view of psychology, it is intuitive to assume that behavioural preferences are external manifestations of internal thinking styles of individuals or groups, and the corollary is that thinking styles preceed or determine behavioural preferences. Maccoby [3] applied the context of cognitive development theory to discern four types of management thinking styles - analyser, energiser, synthesiser, and humaniser. He further argued that successful “information age organisations” require managers with higher preference for synthesising and humanising thinking styles. In a more rogorous approach, Herrmann’s Dominance Instrument [4] articulates twenty thinking styles, classifying them into four quadrants of the Whole Brain Model. Herrmann’s approach as summarised in Figure 2 shows a proforma of “200,000 profiles of the mentality of 200 representative occupations”.

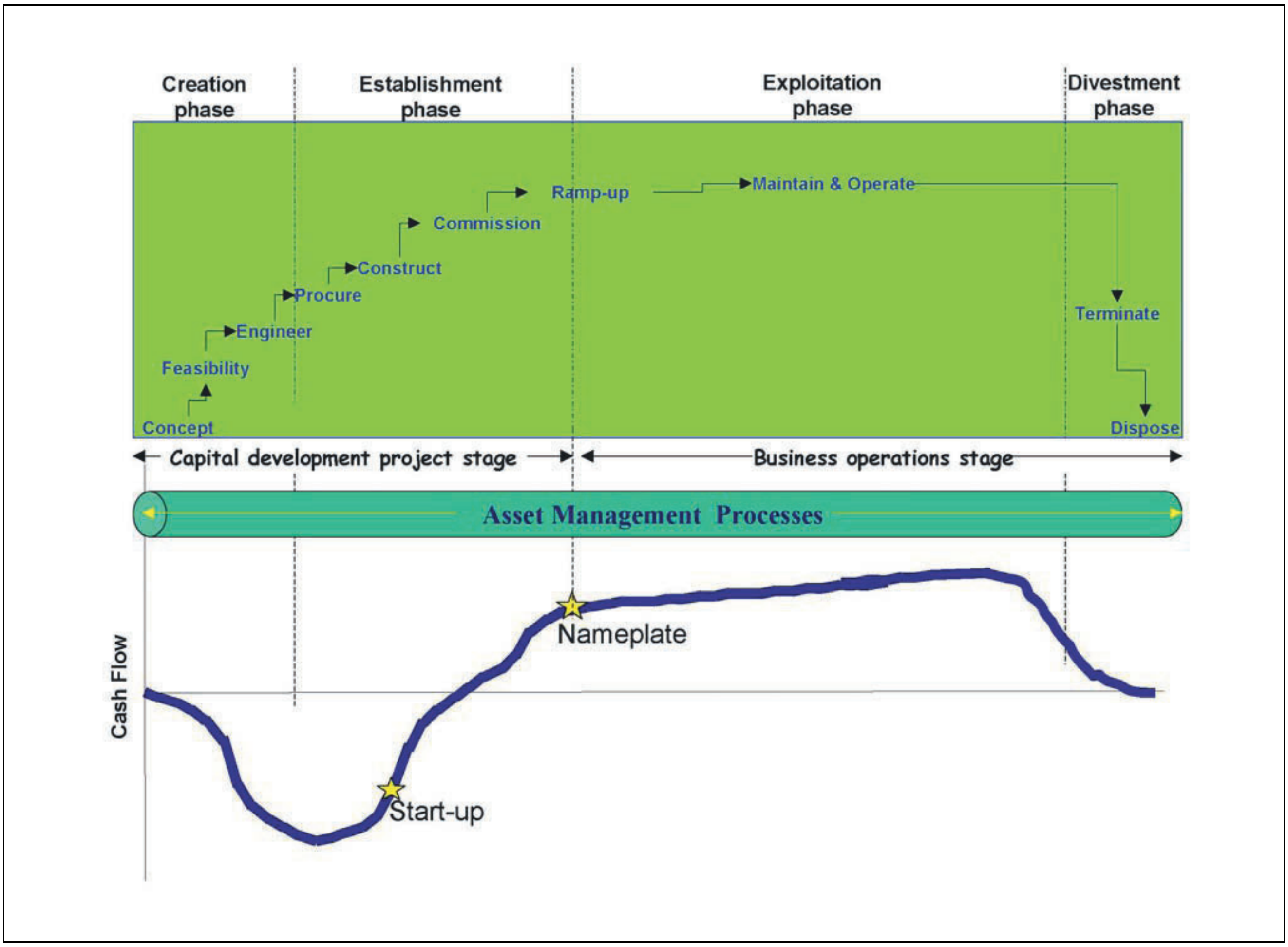


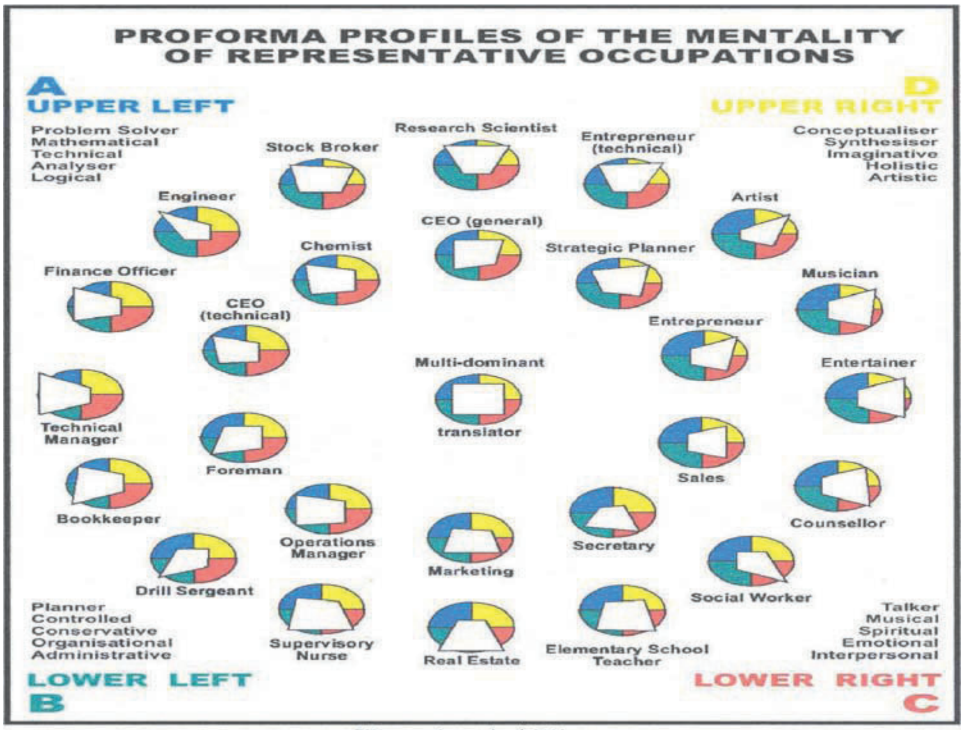
Figure 1. Life-Cycle of Built Environment Physical Asset

The question that arises from Figure 2 is - what should be the profile for the *Engineering Asset Manager*? What thinking styles should predominate so that appropriate behavioural preferences manifest into attitudes which, when combined with knowledge and skills, provide for effective management of physical assets within the new paradigm of innovation economy? The paradox seems to be the tension between conventional thinking styles of engineering professionals and the new requirement for behavioural preferences that encompass a much wider range of cognitive and mental processing modes. The innovation economy paradigm is characterised by continuous acquisition of new skills, with the corresponding rapid conversion and application of ideas and knowhow towards generating the value profile defined by multifarious stakeholders. Perhaps, the perplexing issue is that the Engineering Asset Manager has to rapidly adapt and adopt new mental processing modes necessary for managing the respective life-cycle phases/stages and complex organisational forms associated with the physical asset base within the vagaries of the innovation economy paradigm.

3 RESEARCH

3.1 Survey

The presentation here is an attempt to identify the dominant thinking styles for engineering asset management within the context of the innovation economy paradigm. Whilst on a speaking tour of Australia, the author requested that attendees complete a one-page questionnaire at the start of the meeting. The audience mostly consisted of practising engineers with primary responsibility for maintenance of physical assets. Considering that the engineering professional of the innovation generation is a knowledge worker, respondents were asked to rank the twenty thinking styles described in the Herrmann Brain Dominance Instrument (HBDI) on a five-point scale, ranging from 'not important (score = 1)' to 'extremely important (score = 5)'. All respondents completed the questionnaire within five minutes. The results are presented here in a manner consistent with the four-quadrant HBDI delineation of whole brain model and dominant thinking styles. For brevity, the focus is on the summarised opinions expressed in the feedback, bearing in mind that the data is non-probablistic, that is, the respondents do not form a statistically representative sample of the entire population of engineering professionals.



©Herrmann International, 2000

Figure 2. Thinking Style Profiles of Representative Occupations

3.2 Results

A total of 190 respondents participated in the survey. Table 1 shows the number of respondents who indicated a ranking for each of the twenty thinking styles. Questionnaire feedback in which a ranking was not indicated for more than fifteen thinking styles was regarded as incomplete and not included in the data set.

Table 1: Number of respondents indicating a ranking for Thinking Styles

HBDI Whole Brain Model Delineation of Thinking Styles	HBDI Thinking Styles	Number of respondents indicating a ranking for each Thinking Style
Quadrant D, Upper Right, Cerebral	Imaginative	187
	Synthesiser	183
	Artistic	184
	Holistic	181
	Conceptualizer	183
Quadrant C, Lower Right, Limbic	Interpersonal	185
	Emotional	186
	Musical	185
	Spiritual	185
	Talker	188
Quadrant B, Lower Left, Limbic	Controlled	185
	Conservative	186
	Planner	187
	Organisational	186
	Administrative	186
Quadrant A, Upper Left, Cerebral	Logical	186
	Analyser	186
	Mathematical	186
	Technical	184
	Problem Solver	187

The relative importance of thinking styles as indicated by the respondents is shown in Figure 3 in a manner consistent with the HBDI delineation according to the Whole Brain Model. This picture shows that a discernible number of respondents did not attach any importance to artistic (9.7%), emotional (10.7%), conservative (21.5%), spiritual (33%) and musical (46%) thinking styles. This could be interpreted to mean that as much as 46% of respondents suggest that musical thinking style should not influence the behavioural preference of the *Engineering Asset Manager* of the innovation dispensation.

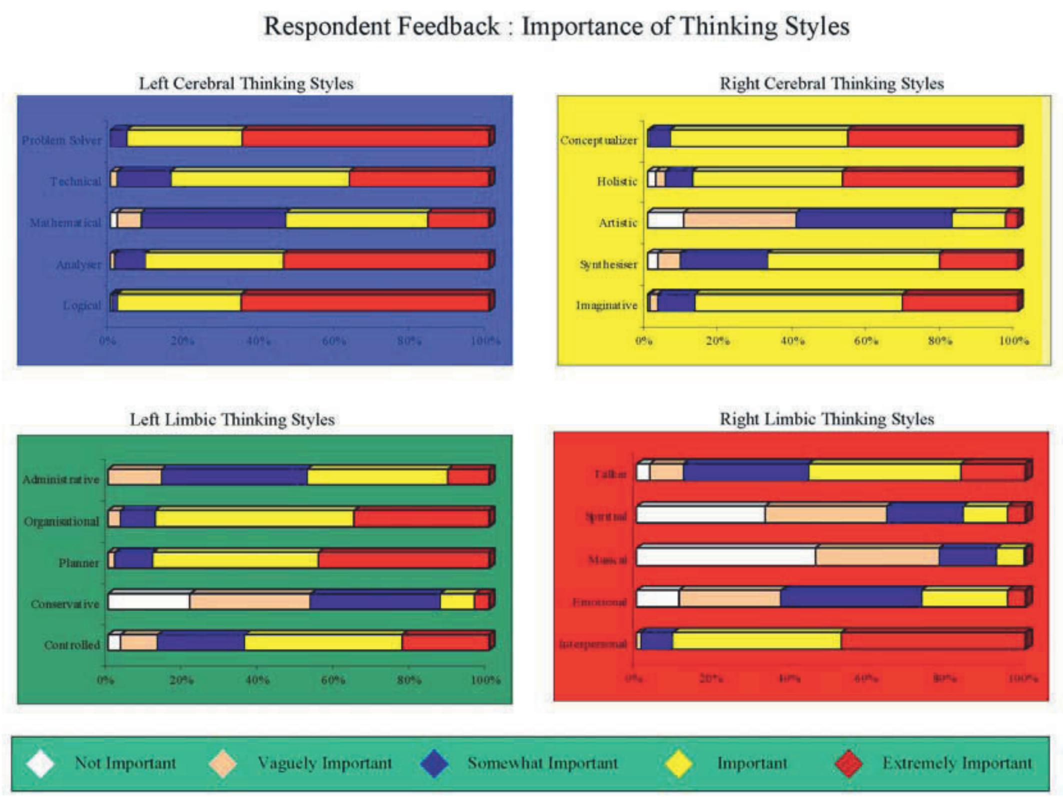


Figure 3. Ranking of Thinking Styles

The top ten thinking styles ranked as extremely important by the respondents is shown in Figure 4 to include:

i. **Logical** – Nearly 66% of respondents indicate that the 'ability to reason deductively from what has gone before' is extremely important.

ii. **Problem solving** – Again, 65% of respondents indicate that the 'ability to find solutions to difficult problems by reasoning' is extremely important.

iii. **Analysing** – 54% of respondents indicate that the 'ability to break up ideas into parts and examining them to see how they fit together' is extremely important.

iv. **Interpersonal** – Nearly 48% of respondents indicate that the 'ability to easily develop and maintain meaningful and pleasant relationships with many different kinds of people' is extremely important.

v. **Holistic** – About 47% of respondents indicate that the 'ability to perceive and understand the 'big picture' without dwelling on individual elements' is extremely important.

vi. **Conceptualisng** – 46% of respondents indicate that the 'ability to conceive thoughts and ideas, to generalise abstract ideas from specific instances' is extremely important.

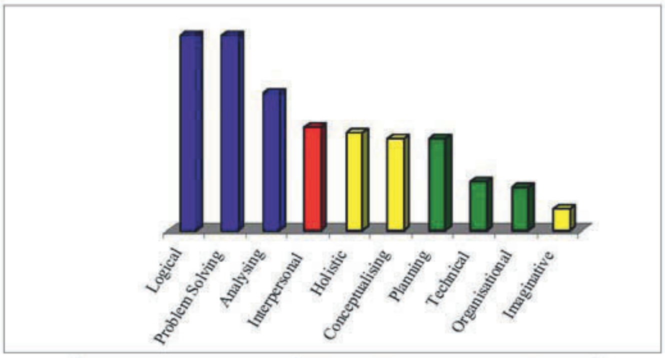
vii. **Planning** – 45% of respondents indicate that the 'ability to formulate methods to achieve a desired end in advance of taking actions to implement' is extremely important.

viii. **Technical** – 37% of respondents indicate that the 'ability to understand and apply engineering and scientific knowledge' is extremely important.

ix. **Organisational** – 35% of respondents indicate that the 'ability to arrange people, concepts, ideas, etc into coherent relationships with each other' is extremely important.

x. **Imaginative** – 31% of respondents indicate that the ‘ability to form mental images of things not immediately available to the senses or never wholly perceived in reality, ability to confront and deal with a problem in a new way’ is extremely important.

Respondent Feedback : Top Ten Extremely Important Thinking Styles



Thinking Style	% of Respondents indicating Extremely Important
Logical	65.5%
Problem Solving	65.2%
Analysing	54.3%
Interpersonal	47.5%
Holistic	47.2%
Conceptualisng	46.0%
Planning	45.0%
Technical	37.0%
Organisational	35.4%
Imaginative	31.0%

Figure 4. Top Ten “Extremely Important” Thinking Styles

4 SUMMARY

Six of the top ten extremely important thinking styles belong to the left-brain quadrants, hence validating the HBDI proforma profile of engineering and related occupations. This is not surprising since all the respondents had engineering background and technical orientation. What is more interesting is the fact that four of the top ten extremely important thinking styles belong to the right-brain quadrants, with three in the upper right mental processing mode. This result also supports the increased shift in emphasis towards behavioural preferences generally refered to as ‘soft’ skills. In comparing the respondent feedback with the HBDI proforma profiles of occupations shown earlier in Figure 2, the suggestion is that the profile for the *Engineering Asset Manager* should lie between that shown for technical chief executive office (CEO) and chemist. The ramification is that right-brained mental processing modes should be combined with scientific knowledge and technical skills to effectively manage the various life-cycle phases, stages, and activities associated with built environment assets.

The essence of managing physical assets that form our built environment is to provide benefits to satisfy the continuum of constraints imposed by rapidly changing business strategy, economy, ergonomics, operational and technical integrity, and regulatory compliance. Innovative approaches to enhance and sustain the profile of values required from these assets demands a shift in the behavioural preferences or attitudes of engineering professionals in asset management occupations. This implies that engineering professionals in asset management occupations need to adapt to new thinking styles, and adopt effective cognitive and mental processing modes. Whilst assuming that thinking styles manifest in attitudes, this paper has described the results of a 2005 survey of 190 practicing engineers to ascertain what thinking styles should determine behavioural preferences of engineering-oriented managers of built environment assets of the innovation generation. The study confirms other results from cognitive theory and psychology, highlighting the top ten thinking styles as ranked by survey respondents. The paper is intended to provide a strategic view of Engineering Asset Management within the context of innovation, with particular focus on behavioural alignment towards the modern paradigm of the knowledge and learning economy.

5 REFERENCES

[1] Amadi-Echendu J. (2005) Physical asset management requires a paradigm shift from maintenance: MESA Tour 2005 and ICOMS 2005 Keynote address. Adelaide, Brisbane, Melbourne, Newcastle, Perth, Sidney, Wollongong and Hobart. 30 May – 10 June 2005.

[2] Amadi-Echendu J. (2003) Developing operational capability during major capital projects: Proceedings of International Conference on Maintenance and Asset Management (ICAMM), Pretoria, South Africa. Oct 2003.

[3] Maccoby M. (1994). From Analyzer to Humanizer: Raising the Level of Management Thinking. Research Technology Management, Vol. 37 No. 5, pp. 57-59.

[4] Herrmann N. (1996). The Whole Brain Business Book, McGraw-Hill ISBN: 0070284628

6 ACKNOWLEDGMENTS

The author wishes to acknowledge sponsorship received from University of Pretoria, ESKOM South Africa, and Maintenance Engineering Society of Australia for the MESA 2005 Eminent Speaker Tour during which the survey was conducted.

A CASE STUDY ON MAINTENANCE AND RSA PUBLIC SERVICE DELIVERY

G Stockwell

Masters in Project Management, Department of Engineering and Technology Management, University of Pretoria, Pretoria 0002, South Africa

Abstract: Public sector service delivery is vital in ensuring that priorities determined by government are met. Maintenance is a vital function in the management of public sector physical assets. The goal is not only to ensure that such assets operate as intended but also, that they remain in good condition and provide the capacity for effective and efficient service delivery.

In order to address a historical backlog in the maintenance of public sector assets, the South African National Department of Public Works has implemented a Repair and Maintenance Programme (RAMP). This paper identifies five key implementation factors of RAMP and their influence on efficient and effective service delivery, as viewed from a survey of the respective custodians of the public sector assets. The paper examines the performance of public sector service delivery and provides a comparison pre- and post-implementation of RAMP. The findings of the research is important to government policy makers with regard to the effectiveness of RAMP.

Key Words: Physical Asset Management, Public Sector Assets, Repair and Maintenance, Customer Satisfaction.

1 BACKGROUND

1.1 Introduction

The South African National Department of Public Works, in partnership with other National Government Departments, has been tasked with promoting commercial attitudes within the public sector with regard to efficiency of service delivery programmes. Service delivery is of key importance in ensuring that priorities determined by the National Government are met. These priorities relate to housing, water, sanitation, health, security, social welfare, education and poverty relief, to mention but a few. The South African National Department of Public Works plays a crucial role in facilitating delivery by other National Government Departments through the provision and management of public sector infrastructure.

A vital part of the management of public sector infrastructure is continuous maintenance to ensure that public sector infrastructure is conducive and capable of ensuring effective and efficient service delivery at all times.

Since the first democratic elections took place in South Africa in 1994 the South African Government has set in place many policies to increase inter-continental commerce within Africa. This has resulted in a vast increase in traffic flow through South African Border Control Ports of Entry. Where previously emphasis was placed on Border Control Immigration Enforcement weight was shifted to enforcement of Customs bringing about a need for increase efficiency of the South African Revenue Services: Customs together with increase numbers of personnel. The increased number of personnel places additional load on the infrastructure and thus more efficient maintenance is required.

Statistics have shown that the prison population in South Africa has increased drastically over the past 20 years. Research conducted by Oppler indicated an increase in the prison population from 109,000 in 1984 to 146,000 in 1997 and further indicated a projected increase to 156,000 in the years 1998 and 1999 (1998, pp.2-3). Oppler stated that the increase in prison population was in line with international trends and that the same phenomenon was evident in the UK and US. The Times accounted that the increases prison population could be attribute to sentencing reforms rather than an increase in crime (1996). Despite the vast increase in the prison population over the above period insufficient additional prisons (additional prisons with a capacity of approximately 10,000 were constructed) have been built and as a result prisons are over-populated resulting in increased maintenance strain on the facilities.

The above two paragraphs serve as examples of the amplified requirements of South African public sector infrastructure over the past 20 years.

Currently the repair backlog of public sector infrastructure in South Africa is estimated at R12.5 billion. Faced with the challenge of reducing the RSA public sector repair backlog over a period of ten years as well as promoting commercial attitudes with regard to service delivery, the South African National Department of Public Works developed the Repair and Maintenance Programme (RAMP) methodology in January 2000. The Repair and Maintenance Programme (RAMP) is a multi-disciplinary programme aimed at repairing public sector infrastructure to a functional condition and maintaining such infrastructure in a functional condition in order to ensure the required service delivery by other National Government Departments.

Since the RAMP was initiated in 1999 and pilot projects implemented for the Department of Correctional Services the programme has escalated to include over 199 facilities for 10 National Government Departments.

To date 615 Professional Service Providers and 390 Contractors have been appointed. Status quo reports to the value of R 5.9 billion have been developed. Expenditure to date on the RAMP amounts to R 3.0 billion.

1.2 Research Problem

The RAMP is currently in the fifth year of implementation and to date the level of service delivery that the RAMP offers to the User Clients/National Government Departments has not been formally assessed, evaluated or quantified.

The expenditure achieved through the implementation of the RAMP for the 2004/2005 financial year amounted to R 1.2 billion. As future planning the National Department of Public Works has developed a holistic maintenance plan that indicates a required R 30.8 billion over a period of 10 financial years.

From the above paragraph it is evident that the RAMP is responsible for a large portion of the National Department of Public Works' yearly budget allocation and expenditure. The findings of this study will be of importance to the South African National Department of Public Works policy makers in order to assess whether the RAMP offers the level of service delivery expected by the National Government Departments and thus assess whether the RAMP offers "value for money" and should be extended in future.

In order to assess the service delivery offered by the RAMP it is firstly required to gauge the effectiveness of the RAMP compared to the previous maintenance methodologies employed namely; (1) day to day maintenance, (2) year/term contracts and (3) own resources. Given the results of the above, the level of improved service delivery post implementation of the RAMP will be assess in terms of a pre- and post- RAMP appraisal.

The performance based maintenance methodology put into action as part of the RAMP relies on 5 key implementation factors namely; (1) monthly site progress meetings, (2) monthly maintenance scoring, (3) logging of breakdowns, (4) feedback of breakdowns repaired and (5) performance based payment. The contribution of each of the factors toward service delivery as well as the ability of User Clients to influence service delivery through their involvement in the 5 factors will be assessed and presented. This will aid the Department of Public Works and User Clients to place emphasis on the correct aspects in order to optimise service delivery.

In order to assess the service delivery offered it is firstly required to establish what factors are of importance to the User Client in the quantification of service delivery. Through cooperation with the User Clients and the Department of Public Works the major factors which are used by the various Government Departments to quantify service delivery will be identified. This will aid the Department of Public Works in future assessment and quantification of service delivery offered to the User Clients.

2 THEORETICAL FRAMEWORK

Literature advocates that (1) Maintenance, (2) Commercialisation of the Public Sector and (3) Outsourcing are factors which significantly contribute toward service delivery.

Coetzee reports that maintenance is the fastest growing discipline in the world (1997). "In the last decade, industrial firms in different industries are increasingly tempted to follow a" life-cycle strategy (construction, development, maintenance etc.), "thus offering total need fulfilment to customers through product/service bundles" (Stremersch, 2000, p.1). Stremerch *et al.* concluded that the reputation of the Contractor and the level of detail and information provided regarding maintenance activities are the two most prominent factors for service delivery in full-service maintenance contracts (2000, pp.9-10).

Brown *et al.* indicated "There is an international trend to contestability and marketisation in the delivery of public services" (2000, p.206). Commercialisation can be defined as "a compromise between traditional style of public administration and marketisation in the form of privatisation, contracting-out and contestability" (2000, pp.206-207). In their research Brown *et al.* explore how the commercialisation of government services result in "greater efficiency, higher quality of service, a clearer focus on customers and better value for money" while Stewart and Walsh indicate the strongest justification for introducing private market-orientated delivery systems are usually based on the need to reduce government expenditure, improve cost efficiency and provide more accountable and transparent delivery systems (1992, pp.97-110). Brown *et al.* also identifies the capacity of government departments to deliver services as a driving factor for commercialisation. One of the major problems identified with public service delivery has been identified by Brown as "fragmentation and the lack of appropriately coordinated government services" (2002, p.439).

Teresko reported that in 1995 Peter Drucker predicted, "in 10 to 15 years organisations may be outsourcing all work that is 'support' rather than revenue producing, and all activities that do not offer career opportunities into senior management" (1999, p.38-45). Since outsourcing was first identified as a business strategy 16 years ago outsourcing has gained acceptance in the marketplace and a growing trend established amongst international organisations and corporations to outsource their services, day-to-day activities, maintenance requirements amongst others (Mullin, 1996, pp.28-36). Outsourcing has gained respect in the market place as “top-performing companies outsource to perform even better” (Teresko, 1999, p.45).

The foremost factors listed for the outsourcing of services was cost-effectiveness and improved customer satisfactions (service delivery). According to Sam Amore and others, in considering outsourcing one should “Evaluate the concept from the standpoint of what the value-added is as opposed to what the reduction in cost is going to be” (Chalos *et al.*, 1998, pp.901-919).

Clarke reports that 92% of companies who outsource listed cost as the driving factor. However Clarke says that the underlying management philosophy, the notion of “core competency” is just as important as the cost aspect (2004, pp.33-35). According to Quinn *et al.* possibly the greatest leverage obtained through outsourcing is that of specialised professional capabilities that would be prohibitively expensive or even impossible to duplicate internally (1994, pp.43-56).

Mullin however indicates that various outsourced contracts are not successful as a result of “many consultants, outsourcing suppliers and companies that outsource agree that businesses have virtually abdicated the management function in outsourcing contracts, creating and environment where suppliers manage the relationship” (1996, pp.29-36). Goldman says that the lack of performance expectations, measurement criteria, procedures for amendments and lack of continuity in personnel involved are the major factors which plague outsourcing (Mullin, 1996, p.29).

3 RESEARCH DESIGN AND METHODOLOGY

3.1 Research Strategy

The research population includes the Department of Public Works (as service provider) and the various User Departments for whom the RAMP is currently being implemented. Data gathering was divided into three hierarchical levels; (1) Department of Public Works Head Office, (2) User Client Head Office and (3) User Client Local Representatives. Separate questionnaires were administered at the three hierarchical levels. Questionnaires were administered by means of a call centre in order to increase the response rate and to further eliminate some potential bias.

Questionnaires at levels 1 and 2 were aimed at obtaining high level information regarding the factors used for quantification of service delivery, input of the RAMP toward a commercialised mindset and the role of the 5 key implementation factors toward service delivery. An evaluation of the success of the RAMP compared to previous maintenance methodologies was conducted. The questionnaires administered at level 3 focussed specifically on an evaluation of service delivery pre- and post- implementation of the RAMP as well as the contribution of the 5 key implementation factors toward service delivery.

For the purposes of the research at hand the 3 independent User Departments with the highest number of facilities where the RAMP is currently being implemented were selected as the sampling frame. The three User Departments selected are; (1) Department of Correctional Services, (2) Department of Justice and (3) Border Control Ports of Entry.

3.2 Data Analysis

Data was collected and analysed per User Department. The data presented here illustrates an overview of the data gathered for the three User Client Departments.

3.2.1 Quantification of Service Delivery

The quantification of service delivery was only assessed at User Client Head Office level as it is at Head Office level that service level agreements are agreed upon. The research highlighted 6 key factors that could be used for quantification of service delivery namely; (1) Expenditure vs. Budget, (2) Number of contracts completed successfully vs. the number of contract not completed successfully, (3) Contract completed on time vs. contract not completed on time, (4) Contract cost, (5) Continuity of work and role-players and (6) Time from planning instruction to site handover. The User Client Head Office level rated the 4 factors most and equally important for the quantification of Service Delivery as being time for completion, contract cost, continuity of work and time from planning instruction to site handover. Expenditure vs. budget is least important for the quantification of service delivery.

The User Department listed the following factors which should additionally be considered in the quantification of service delivery; (1) Personnel wellness, safety and health and happiness, (2) User Client Local Representative satisfaction, (3) Shorter planning and implementation times, (4) Quality of work and (5) Follow up service.

3.2.2 Commercialisation

By definition the RAMP can be classified as a programme which contributes toward the commercialisation of the public sector. All User Client Department confirmed that the RAMP contributes toward a commercialised mindset within the RSA Public Sector.

3.2.3 Effectiveness of Previous Maintenance Methodologies

The vast majority of the User Clients at Head Office and Local Representative Levels rated the effectiveness of previous maintenance methodologies as less effective than the RAMP as illustrated by Figure 1.

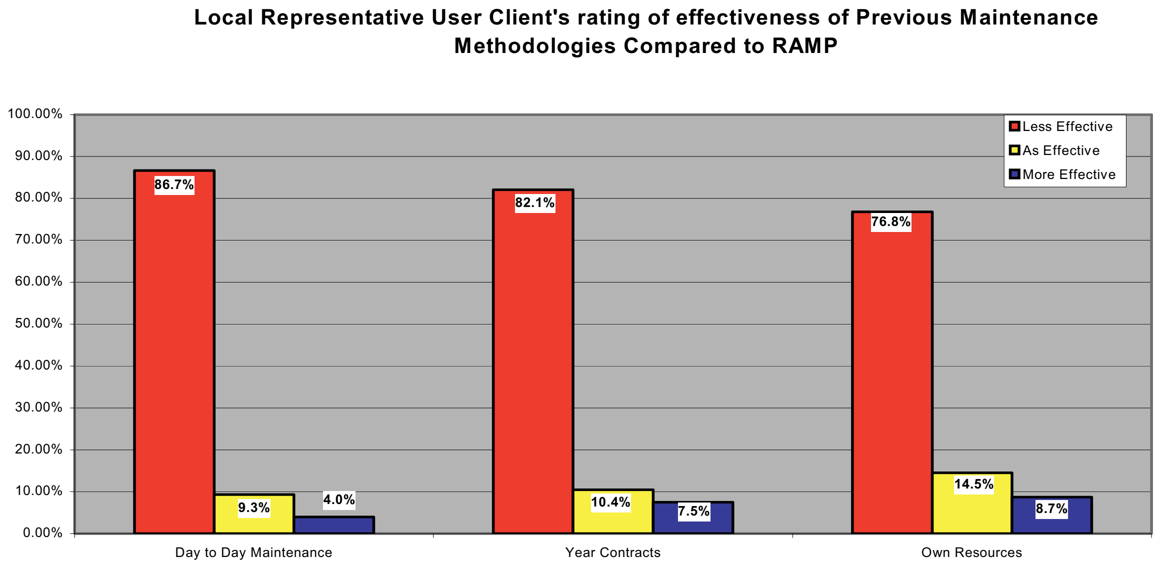


Figure 1: Effectiveness of previous maintenance methodologies compared to RAMP

3.2.4 Contribution of the 5 Key implementing factors of the Performance Based Maintenance Methodology toward Service Delivery

The 5 key implementation factors employed as part of the performance based maintenance system are; (1) Monthly Progress Meetings, (2) Monthly Maintenance Scoring, (3) Logging of breakdowns, (4) Feedback of breakdowns repaired and (5) Performance based payment. The contribution of each of the above factors was rated by the User Client in terms of major- (5 points), strong- (4 points), medium- (3 points), minor- (2 points) and no- influence (1 point).

The Head Office User Clients rated feedback of breakdowns repaired, logging of breakdowns and performance based payment as being the major contributing factors toward service delivery. The Local Representative User Clients identified monthly progress meetings, logging of breakdowns and the feedback of breakdowns repaired as the major contributors toward service delivery.

3.2.5 Improvement of Service Delivery through User Client's Involvement in the 5 Key implementing factors of the Performance Based Maintenance Methodology

The Head Office User Client views the User Client's involvement in monthly progress meetings and logging of breakdowns as paramount to their contribution toward improving service delivery followed by their involvement in the monthly maintenance scoring and feedback on breakdowns repaired.

The Local Representative User Client views their involvement in monthly progress meetings, logging of breakdowns and feedback on breakdowns repaired as major contributors toward service delivery.

3.2.6 10 Core Business Requirements: Before & After Ramp And Impact Assessment

Each User Department was requested to identify their 10 core business requirements. The service delivery prior to and post implementation of the RAMP with specific focus on their 10 core business requirement was then rated by the User Clients at Local level. The impact of the improved level of service delivery was assessed in terms of the User Clients increased ability to perform their core business.

Figure 2 indicates a marked overall improvement post implementation of the RAMP for the 10 core business requirements for the Border Control Ports of Entry.

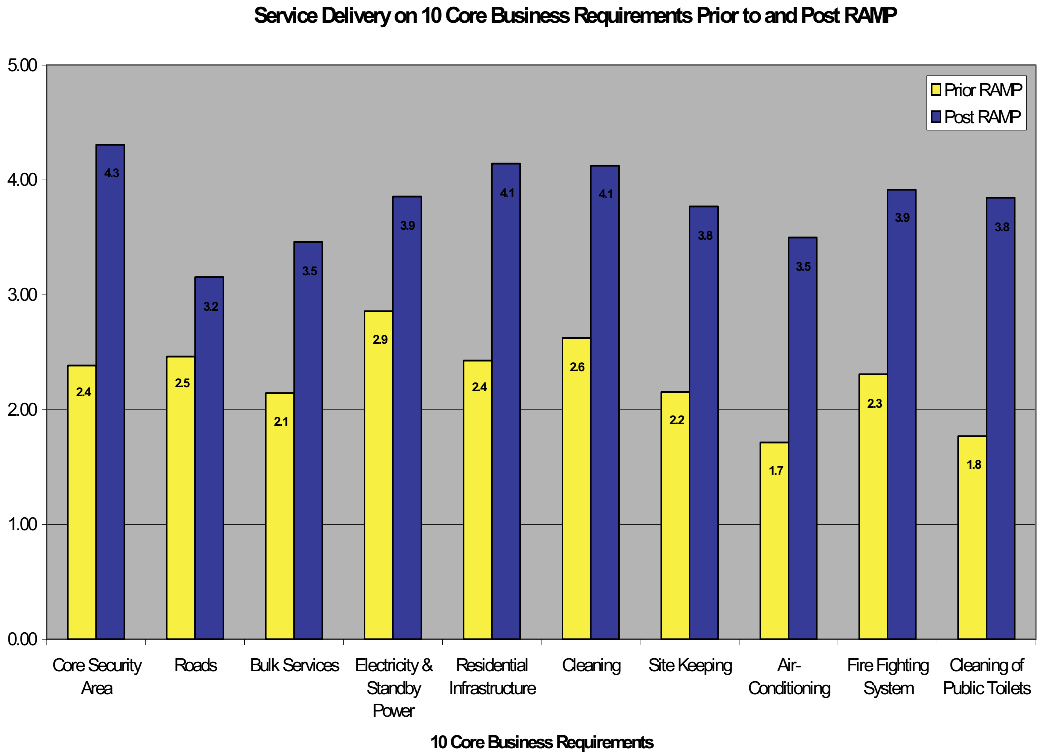


Figure 2: Service Delivery prior to and post the RAMP with reference to the 10 Core Business Requirements

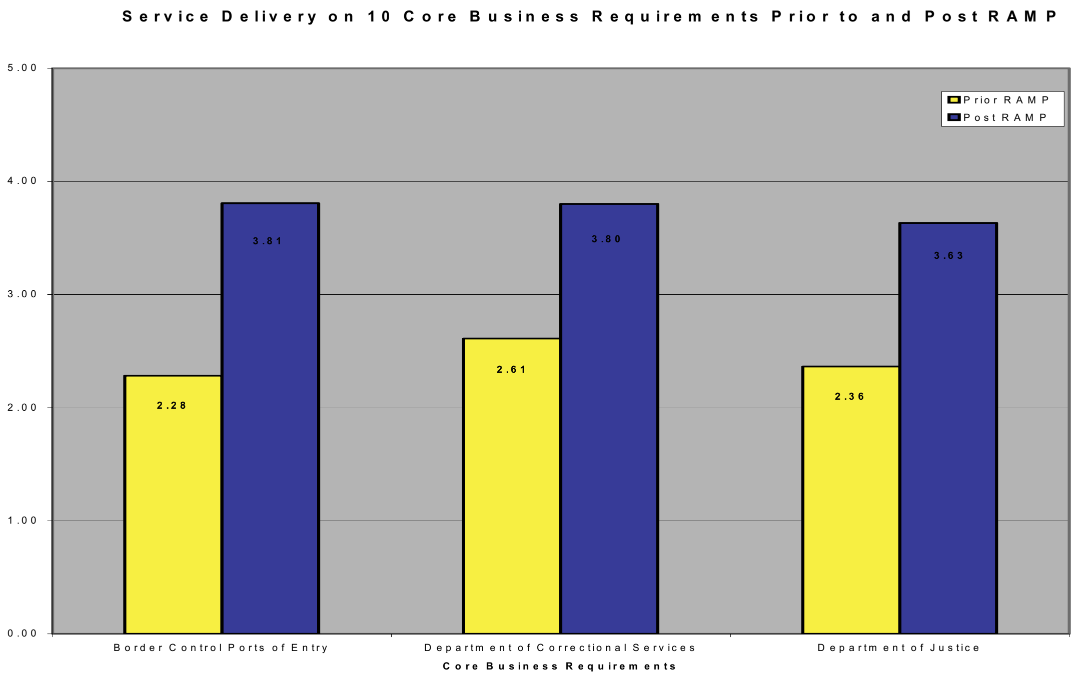


Figure 3: Service Delivery prior to and post the RAMP with reference to the 10 Core Business Requirements.

Figure 3 indicates the average service delivery levels pre- and post- RAMP for the 10 core business requirements of each User Client

4 CONCLUSIONS

4.1 Research Results

The results clearly indicate that maintenance, within the RAMP, contributes toward service delivery within the RSA Public Sector. A clear improvement in service delivery pre- and post- implementation of the RAMP is noticeable. The study further ascertained that the User Client has a substantially improved ability to perform their core business functions as a result of the improved service delivery offered by the RAMP.

The study indicated the RAMP proved to be substantially more effective than any of the previously implemented maintenance methodologies. A far greater structured approach to maintenance is achieved through the implementation of the RAMP compared to the previous maintenance methodologies thus offering greater utility of the facilities.

The implementation of the RAMP has alleviated the burden of maintenance on the User Client (as compared with previous maintenance methodologies) as the burden is now born by the Department of Public Works. This provides the User Client with the opportunity to apply greater focus on performing their core business responsibilities.

The research further yielded factors which can be used for the quantification of public sector service delivery by the User Client Departments and the implementing Department (I.e. National Department of Public Works). These factors may be employed by the Department of Public Works in future service level agreements with other National Government Departments.

The 5 key implementing factors of the RAMP were assessed and found to contribute greatly toward RSA public sector service delivery within a maintenance programme. The two most important factors to both Head Office and Local Representative level were found to be the logging of breakdowns and feedback on breakdowns repaired. Local Representative User Clients also viewed monthly maintenance control meetings as a major contributor toward service delivery.

4.2 Contribution to Practise

The research clearly identified the need for maintenance as a service delivery characteristic within the public sector. The Department of Public Works is in the process of implementing a Service Delivery Improvement Programme (SDIP). The research conducted will aid the Department of Public Works in assessment of the current contribution of maintenance toward service delivery and possible methods for improvement of the current maintenance methodologies to achieved improved public sector service delivery.

4.3 Recommendations

The research has indicated that the User Client's involvement in the key implementing factors of the performance based maintenance methodology is paramount to improved service delivery. It is this essential the User Client is properly informed and involved with the implementation of the keys aspect of performance based maintenance methodology in order to ensure that service delivery to the public sector is optimised.

It is thus suggested that the Department of Public Works imposes more stringent implementation, integration and coordination criteria between the Repair and Maintenance Programme management team and User Client. The User Client is to be actively involvement in the implementation of the 5 key implementing factors of the performance based maintenance methodology.

4.4 Concluding Remarks

The Repair and Maintenance Programme implements a structured approach to maintenance thus providing the User Client with more effective use of public facilities and in so doing lessens the burden of maintenance on the User Client. User Clients are thus able to focus more on performing their core business competencies.

The research has indicated greater efficiency in terms of the maintainability of public sector assets. Although various facility specific problems were identified during the course of the study it can be concluded that the RAMP provides the User Client with greatly improved service delivery in terms of maintenance.

The research has indicated that greater emphasis should be placed on User Client involvement and high-level management within the RAMP.

The RAMP offers more efficient and effective performance compared to any previous maintenance methodologies implemented within the public sector in South Africa. However by applying greater focus in the areas identified by this study public sector service delivery can be further improved.

5 REFERENCES:

[1] Brown, K., Ryan, N. & Parker, R., New modes of service delivery in the public sector. Commercialising government services. The International Journal of Public Sector Management, Vol. 13 No. 3, 2000, pp. 206-221, MCB University Press.

[2] Brown, K., (2002) The Government Service Delivery Project, Public Management review, Vol. 4, Issue 4, pp. 439-459

[3] Clarke, R., The real cost of outsourcing, America's Network, May1 2004, pp. 33-35

[4] Coetzee, L.J., 1997. Maintenance. Maintenance Publishers (Pty) Ltd: South Africa

[5] Mullin, R., Managing the outsourced enterprise, Journal of Business Strategy, July/August 1996, pp. 29-36

[6] Oppler, S., Assessing the state of South African Prisons, South African Security Review, Vol. 7 No. 4, 1998

[7] Quinn, J.B. and Hilmer, F.G., Strategic Outsourcing, Sloan Management Review, Summer 1994, Vol. 35 Issue 4, p 43

[8] Quinn, J.B., Strategic Outsourcing, Leveraging Knowledge Capabilities, Sloan Management Review, Summer 1999, Vol. 40 Issue 4, p9

[9] Stremersch, S. Wuyts, S. and Frambach, R.T., The purchasing of full-service contracts: An Exploratory study within the industrial maintenance market

[10] Teresko, J., ERP Outsourcing, E Business, September 1999, pp. 38-45

[11] The Times, 4 April 1996

CRITICAL SUCCESS FACTORS FOR BROWN-FIELD CAPITAL AND RENEWAL PROJECTS

J Warchol [a] and Prof J Amadi-Echendu [b]

[a] *Impala Platinum Ltd – Refineries, cnr East Geduld Rd and Enstra St, Springs 1559, South Africa.*
[b] *Department of Engineering and Technology Management, University of Pretoria, Pretoria 0002, South Africa.*

Abstract: Manufacturing companies operate in a business environment where, due to the factors such as cost and location of the infrastructure, incremental growth and competitive advantage may be a result of in-house construction projects. Competitive advantage may be realised by effective management of the critical success factors of in-house construction projects. These projects often tend to be of a brown-field nature and are characterised by a high degree of uncertainty, predominantly due to the interaction between the project and the existing physical asset base. In this regard, it is vital to understand the factors that influence the success of capital and renewal construction projects in the brown-field context. Although each project has unique features, there are critical success factors which can be customised towards successful outcomes in the brown-field environment. Based on a survey of recent projects, this paper examines critical success factors applicable to brown-field capital and renewal construction projects. The paper also provides an empirical assessment of the degree of managerial focus on the critical success factors and its impact on the outcome of brown-field construction projects.

Key Words: Physical Asset Management, Brown-field Projects, Critical Success Factors.

1 INTRODUCTION

Project Management is a well-structured discipline offering approaches that may be applied to almost any activity or aspect of human life spanning across the cultures, industries and organisations. Project Management evolves all the time in order to accommodate the latest requirements demanded by the new circumstances or by the social or technological developments. It is widely acknowledged that successful companies increase their market share through the successful execution of their projects. The market environment of the 21st century demands high quality of products within an acceptable price range and delivered in the shortest possible time. As a result, organisations must continually improve the quality and the efficiency of their processes as well as the cost effectiveness of their products to remain competitive. In order to survive in such a competitive market, organic growth of an existing asset base is essential. This necessitates a strategically planned and executed portfolio of projects to create economic value and competitive advantage.

Manufacturing companies operate within very specific business dynamics that require a focus on sustained growth. These companies increase their competitiveness through unit cost reduction and compliance with contractual obligations (quality of product and reliability of product delivery) rather than price setting. As a result, there is an ongoing need within these organisations to continually improve their production facilities and processes. While growth provided by completely new plant facilities is not uncommon, the incremental growth of a manufacturing company may be a result of a revolving program of in-house construction projects. Due to the factors such as cost, location and presence of an existing infrastructure, majority of these projects tend to be of a brown-field nature.

Brown field denotes “a land previously built on” (Encarta Dictionary, 2005). Accordingly, *brown field project* may be defined as a project executed within the existing facilities and requiring some demolition or remediation (Corman, 2004). Brown-field construction project is typically executed within, and usually at the mercy of, an existing asset (plant) environment. High degree of uncertainty is a predominant factor influencing project success as unexpected schedule or scope changes due to the plant operating requirements or scope changes are not uncommon. Project Management process must also accommodate the over-riding focus of management on the performance of the asset in use.

By comparison, *green field* project may be defined as a project executed on a virgin land and not affected by the existing production facilities. As a result, the uncertainties associated with the impact of the existing production facilities on the project do not feature in the project’s overall risk profile.

Although project success is critical to any organisation as its survival often depends on it, success is difficult to achieve. By definition, every project is a unique endeavour and is subject to uncertainty. Ideally, project uncertainty should be gradually resolved during the project execution. To facilitate this process, it is essential to establish and exploit factors which are critical to the project success (critical success factors). Due to the uniqueness of all projects, each project may be dependent on a different set of critical success factors. The body of knowledge available, however, indicates that projects executed in a similar environment and having similar goals exhibit a marked similarity in terms of the applicable clusters of critical success factors. As a result, brown-field construction projects should have a specific set of critical success factors which, if exploited, should increase chances of the project success.

Previous experience in the execution of brown-field construction projects indicates that the generous body of knowledge available in the Project Management domain needs to be customised in order to address the practicalities of such projects. Above all, it is necessary to identify the relevant critical success factors applicable to brown-field construction projects and manage these factors on an ongoing basis in order to improve success rate. Unfortunately, little information exists in the public domain regarding the identification of the critical success factors for brown-field construction projects. It is also unknown to what degree a managerial focus on the critical success factors improves the success rate of brown-field construction projects.

The research objective is to identify a set of critical success factors specific to brown-field construction projects and to establish correlation between the management focus and project outcome. This paper attempts to answer the following questions:

- What critical success factors are applicable to brown-field construction projects?
- What is correlation between managerial focus (effort) on the critical success factors and the overall outcome (success) for brown-field construction projects?

Two hypotheses arise:

- Hypothesis 1: A project-specific set of critical success factors exists for brown-field construction projects.
- Hypothesis 2: An increased managerial focus in the areas affected by the critical success factors results in higher rate of success of brown-field construction projects.

2 THEORETICAL BACKGROUND

Project Management is usually tracked back to the precedence network diagramming techniques developed for the Polaris Submarine project in the 1950's (Fondahl, 1987). Today, virtually all construction, product development and engineering efforts use some formal Project Management structure (Cleland & King, 1983). There is seemingly no common definition of Project Management. Oisen (1971) defines Project Management as " ... a collection of tools and techniques ... to direct the use of diverse resources toward the accomplishment of a unique, complex one-time task within time, cost and quality constraints …". Atkinson (1999) points out the link to some project success measures such as cost, time and quality.

“Grow or die” is a well-known business adage that symbolises why many companies have a long-term goal of sustained growth (Bloodgood & Katz, 2004, p.60). This goal necessitates a strategically planned and executed portfolio of successful projects. Since projects are a part of the organisation’s strategic management, defining and assessing project success is therefore a strategic management concept (Shenhar, Dvir, Levy & Maltz, 2001). Building on this view, it is critical to identify success measures, and their respective sensitivity to project success. Cost, time and quality (the Iron Triangle) have become inextricably linked over the last 50 years with measuring the success of a project. Atkinson (1999, p.339) proposed two-stage framework viz.: criteria for success in the delivery stage ("doing it right") and criteria for success in the post-delivery stage ("getting it right"). Cost, time and quality fall into the former group while impact on customer, business success resulting from the project and how well the project prepared the organisation for the future fall into the latter part of the framework. Other researchers generally support this line of thinking. Shenhar et al (2001) identified four major distinct success dimensions: project efficiency, impact on the customer, direct business and organisational success and preparing for the future. White & Fortune (2002) conducted a survey which indicated that even though most of the respondents judged the success of their projects by means of the criteria of time, budget and specification, they were also concerned with the impact of the project on the organisation involved. Some studies suggested adding a new element to the notion of project success - client satisfaction (Lim & Mohamed, 1999 and DeCotiis and Dyer, 1979). Baker & Fisher (1988) went one step further to include the level of satisfaction of four different stakeholders: the customer, the developer, the project team and the end user. Some researchers investigated the issue in the reverse manner, by researching the causes of project failure. Resulting from this approach, Pinto & Mantel (1990) identified three aspects of project performance as benchmarks for measuring the success or failure of a project: the implementation process, the perceived value of the project and the client satisfaction with the result. Shenhar et al. (2001) and Pinto & Mantel (1990) also indicated that the concept of project success was contingent on the specific project characteristics.

Project success and failure were first introduced by Rubin and Seeling (1967). Shultz, Slevin and Pinto (1987) were among the first to classify the critical factors. Unfortunately, there is little agreement on the casual factors of project success (Pinto & Slevin, 1987). Belassi and Tukel (1996) proposed to group factors into four areas: those related to project, those related to project manager and the team, those related to the organisation and those related to the external environment. These groups are interrelated such that a factor in one group might influence a factor in another group (Dvir, Lipovetsky, Shenhar & Tischler, 1998, Belassi & Tukel, 1996, White & Fortune, 2002 and Westerveld, 2003). Although each project is unique, similar types of projects may exhibit distinctive patterns of factors associated with success or failure (Pinto & Mantel, 1990).

While each project might require a different set of critical success factors, similar projects should have similar groups of critical success factors. It is, therefore, possible to define groups of success factors for any type of project. These groups of factors must thereafter be carefully managed throughout the project lifecycle in order to achieve success.

3 RESEARCH

The first purpose of the study was to identify a cluster of critical success factors affecting the outcomes of brown-field construction projects. The second purpose was to establish the relationship between managerial focus on these factors and project outcome. The set of applicable success factors was established based on the selection and ranking of the most common factors as viewed by a sample project population. Project success measures were established in a similar manner by surveying another population group. Managerial focus on the identified set of critical success factors was evaluated by means of structured interviews conducted with the key personnel of a selected group of projects. The correlation between the managerial focus on the critical success factors and the project outcome was thereafter established through the analysis of the collected data. Our study concentrated on three distinctive areas, each of them requiring its own process.

3.1 Critical success factors

Twenty factors were selected based on literature review. A survey questionnaire was used to test the perceptions of the project community with respect to the importance of each identified factor. Respondents were requested to asses the impact of each factor on both green-field and brown-field project environment as well as to assess the importance of each factor for the success of brown-field construction projects. All success factors were subsequently ranked in terms of their importance and the five highest ranked factors which were considered to have more impact on brown-field projects as compared to green-field projects were selected as the critical success factors for brown-field construction projects.

Table 1 illustrates the selection process for the critical success factors. The impact rank indicates the perceived impact of the success factor on brown-field construction projects while the importance rank indicates a perceived importance of each success factor within the brown-field project environment. The critical success factors are those which have the highest combined level of importance and impact.

Table 1

Identification of the critical success factors

DESCRIPTION OF SUCCESS FACTOR	HIGHER IMPACT ON BROWN-FIELD PROJECT	IMPACT RANK (A)	IMPORTANCE INDICATOR	IMPORTANCE RANK (B)	RANK TOTAL (A + B)	CSF NUMBER
Consultation and communication with client	37	4	144	3	7	CSF 1
Recognising complexity	44	1	127	10	11	CSF 2
Commitment of end user	39	3	130	9	12	CSF 3
Flexible approach to change	34	5	125	12	17	CSF 4
Taking account of external influences	41	2	115	17	19	CSF 5

The relationship between the responses of the three stakeholder groups was investigated to establish the accuracy of the collected data with respect to the importance of each critical success factor. Figure 1 indicates a high correlation of the obtained responses between all sub-groups of the respondents.

The statistics related to identifying the critical success factors are summarised in Table 2.

Table 2

Statistics for CSF identification

	POP TOTAL	POP%	OBS FREQ (O)	EXP FREQ (E) = POP% * N	CHISQ CALC = $(O - E)^2 / E$
Client team	23	41.1%	72	76.39	0.052
Consultants	11	19.6%	38	36.54	0.109
Contractors	22	39.3%	76	73.07	0.123
TOTAL	56	100.0%	186 = N	186 = N	**0.285**

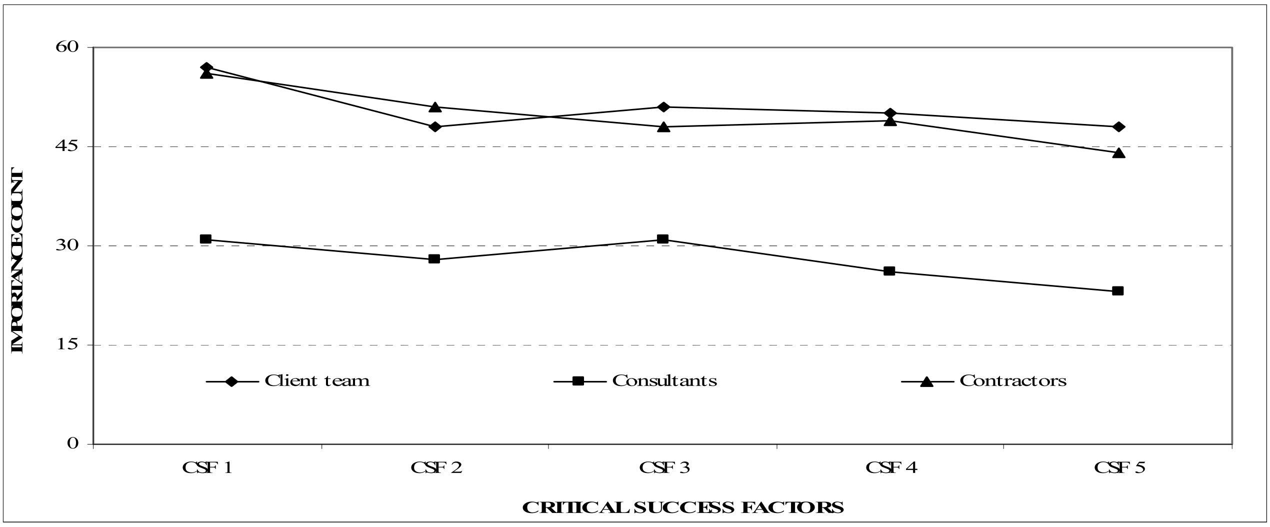


Figure 1. Correlation between respondent sub-groups for the importance of CSF's

3.2 Project success measures

Identification of project success measures for brown-field construction projects commenced with a literature review to establish the most common models for the assessment of the project success. Model proposed by Lim and Mohamed (1999) was selected as we deemed it most suitable for the type of the researched project environment. Ten measures were also selected based on literature review to characterise either the project implementation process or the client satisfaction from the project results. A survey questionnaire was used to test the perceptions of the project community with respect to the importance of each identified factor. Respondents were requested to asses the impact of each factor on both green-field and brown-field project environment as well as to assess the importance of each measure as perceived contribution of each measure towards the project success. All success measures were subsequently ranked in terms of their importance and the five highest ranked measures were selected as the success measures for brown-field construction projects.

Table 3 illustrates the selection process for the success measures. The impact indicator indicates a perceived impact of each success measure on the project, combined for all respondents.

The relationship between the responses of the four stakeholder groups was investigated to establish the accuracy of the collected data with respect to the importance of each success measure. Figure 2 indicates a high correlation of the obtained responses within the project team and the client team but no correlation between the project team personnel and the client team personnel.

3.3 Correlation between the managerial focus on critical success factors and project success

A two-step process was applied to establish the degree of correlation between the managerial focus on critical success factors and project success. Firstly, a suitable database of projects was examined and ten projects were selected to cover most of the variables associated with brown-field construction projects. The following factors were taken into account in the selection process: technological advancement, project size (planned budget), access restrictions (resulting in delays and increased logistical difficulties), project duration (planned duration), impact of project on the exiting asset (plant), impact of the exiting asset (plant) on project and project type. The data was derived from a historical database of projects executed in the period 2002 – 2005. Secondly, the degree of managerial focus on critical success factors was evaluated for each project in the database. At the same time, the success of each project was also evaluated. Finally, the relationship between the two variables was examined.

Respondents were requested to assess managerial focus on each of the five critical success factors for every project in which they participated. Respondents were also requested to rate project success by evaluating each of the five project success measures.

The results obtained from the research were totalised per project - separately for critical success factors and for project success measures. This resulted in an array of x-y parameters describing two variables: critical success factors (independent variable) and project success measures (dependent variable). The findings are represented in Figure 3.

Table 3

Identification of the success measures

MEASURE NUMBER	DESCRIPTION OF SUCCESS MEASURE	IMPACT INDICATOR
SM 1	It solves problem for which it was created	121
SM 2	Customer uses its end product (plant, service etc)	119
SM 3	Customer / user is satisfied with its end product	119
SM 4	It meets operational performance (end product performs designed functions)	117
SM 5	It meets technical specifications (end product conforms to technical specifications)	116

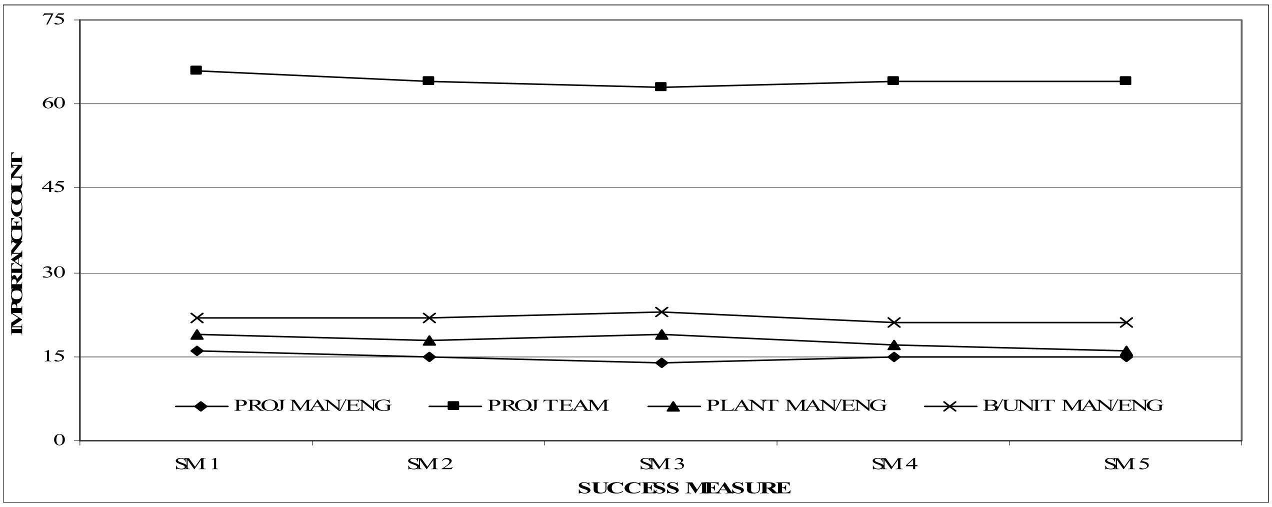


Figure 2. Correlation between respondent sub-groups for the importance of success measures

4 SUMMARY AND RECOMMENDATIONS

Brown-field projects may provide strategic weapon to create economic value and competitive advantage. Although each project may be a unique endeavour, the study supports the idea of clusters of success measures and critical success factors for the brown-field construction environment. In addition, the study points to the conclusion that there is a positive relationship between managerial focus on critical success factors and project outcome.

Test results also indicate a possible perception differences between various project stakeholder groups in two areas: the relative importance of each success measure and the perceived focus of the project team on the critical success factors.

Based on the study results, it is proposed that the success rate of brown-field construction projects may be improved by increased focus on the project – client communication as well as adequate understanding of the complexity associated with this type of projects.

In order to improve the accuracy of the study, it is proposed that further studies be carried out comprising wider study population.

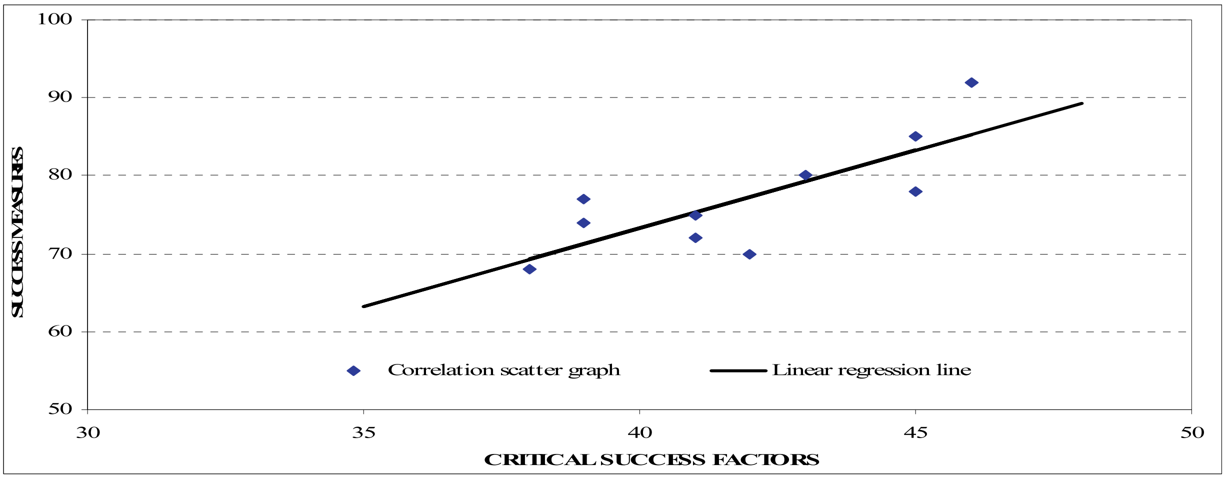


Figure 3. Relationship between brown-field CSF's and success measures

5 REFERENCES

[1] Microsoft Corporation, 2005. *Encarta ® World English Dictionary, North American Edition*. http://encarta.msn.com/dictionary,visited on 8 April 2005.

[2] Corman, T. *Assignment Revamp: The Challenges of a Chemical Plant Upgrade.* In: http://www.maxwideman.com/guests/revamp/intro.htm, visited on 27 December 2004.

[3] Fondahl, J.W. (1987) The History of Modern Project Management. *Project Management Journal*, 28(2), 33-36.

[4] Cleland, D.I., King, W.R. (1983) *Systems Analysis and Project Management.* Mc Graw-Hill, New York. In: Shenhar, A.J. (2001) One Size Does Not Fit All Projects: Exploring Classical Contingency Domains. *Management Science*, 47(3), 394-415.

[5] Oisen, R,P. (1971) Can Project Management Be Defined*? Project Management Quarterly*, 2(1), 12-14. In: Atkinson, R. (1999). Project Management: Cost, Time and Quality, Two Best Guesses and a Phenomenon, It's Time to Accept Other Success Criteria. *International Journal of Project Management*, 17(6), 337-342.

[6] Atkinson, R. (1999) Project Management: Cost, Time and Quality, Two Best Guesses and a Phenomenon, It's Time to Accept Other Success Criteria. *International Journal of Project Management*, 17(6), 337-342.

[7] Bloodgood, J.M., Katz, J.P. (2004) Manufacturing capacity, market share, and competitiveness. *Competitiveness Review*, 14(i1-2), 60-73.

[8] Shenhar. A.J., Dvir, D., Levy, O., Maltz, A.C. (2001) Project Success: A Multidimensional Strategic Concept. *Long Range Planning*, 34, 699-725.

[9] White, D., Fortune, J. (2002) Current Practice in Project Management - an Empirical Study. *International Journal of Project Management*, 20, 1-11.

[10] Lim, C.S., Mohamed, M.Z. (1999) Criteria of Project Success: An Exploratory Examination. *International Journal of Project Management*, 17(4), 243-248.

[11] DeCotiis, T.A., Dyer, L. (1979) Defining and Measuring Project Performance. *Research Management*, 16, 17-22.

[12] Baker, B.N., Fisher, D. (1988) Factors Affecting Project Success. In: Cleland, D.I., King, W.K. (eds) (1988) *Project Management Handbook.* New York, Van Nostrand.

[13] Pinto, J.K., Mantel, S.J.Jr. (1990) The Causes of Project Failure. *IEEE Transactions on Engineering Management*, 37(4), 269-276.

[14] Rubin, I.M., Seeling, W. (1967) Experience as a Factor in the Selection and Performance of Project Managers. *IEEE Transactions on Engineering Management*, 14(3), 131-134.

[15] Shultz, R.L., Slevin, D.P., Pinto, J.K. (1987) Strategy and tactics in a Process Model of Project Implementation. *Interfaces*, 17(3), 34-46.

[16] Pinto, J.K., Slevin, D.P. (1987) Critical Factors in Successful Project Implementation. *IEEE Transactions on Engineering Management*, 34(1), 22-27.

[17] Belassi, W., Tukel, O.I. (1996) A New Framework for Determining Critical Success / Failure Factors in Projects. *International Journal of Project Management*, 14(3), 141-151.

[18] Dvir, D., Lipovetsky, S., Shenhar, A., Tishler, A. (1998) In Search of Project Classification: A Non-universal Approach to Project Success factors. *Research Policy,* 915-935.

[19] Westerveld, E. (2003) The Project Excellence Model®: Linking Success Criteria and Critical Success Factors. *International Journal of Project Management*, 21, 411-418.

CONDITION MONITORING OF LOW SPEED BEARINGS: A COMPARATIVE STUDY OF THE ULTRASOUND TECHNIQUE VERSUS VIBRATION MEASUREMENTS

Yong-Han Kim [a], Andy C C Tan [a], Joseph Mathew a and Bo-Suk Yang [b]

[a] *CRC for Integrated Engineering Asset Management,*
School of Engineering Systems, Queensland University of Technology, 2 George St, QLD4001, Australia.

[b] *School of Mechanical Engineering, Pukyong National University,*
San 100, Yongdang-dong Nam-gu, Busan, 608-739, South Korea.

Abstract: Bearing failure is often attributed to be one of the major causes of breakdown in industrial rotating machines that operate at high and low speeds. This paper presents results of a comparative experimental study on the application of the ultrasound technique for condition monitoring of low speed rolling element bearings and conventional vibration measurements with seeded faults on inner-race defects. The effectiveness of the ultrasound technique is demonstrated through signal processing techniques; use of statistical parameters derived from the time domain, and enveloped spectra in the frequency domain. The results reveal that the ultrasound technique is more effective in detecting low speed bearings failure than that of the vibration measurement.

Key Words: Condition Monitoring, Low Speed Machinery, Rolling Element Bearings, Ultrasound Technique, Vibration Measurements

1 INTRODUCTION

Condition Monitoring (CM) is the process used to determine the operational state and health of a machine for the purpose of detecting potential failures before they turn into functional failures. The CM process consists of periodical or continuous data collection, data analysis, interpretation and diagnosis. Condition Monitoring is an integral part of Predictive Maintenance (PM) which is a widely used maintenance philosophy (also known as Condition-Based Maintenance - CBM). This philosophy is based on scheduling maintenance activities only when a possible functional failure is detected. CBM enhances operational effectiveness and safety through research and development of technologies to accurately and reliably predict the remaining useful life of in-service equipment. CM optimizes equipment readiness while reducing maintenance and staffing requirements. Typical CM techniques include vibration analysis, oil analysis, wear particle analysis, ultrasonic analysis, thermographic analysis and motor current signature analysis [1, 2].

One of the more popular tools in the condition monitoring of rotating machinery is vibration analysis. By measurement and analysis of the vibration in rotating machinery, it is possible to detect typical faults such as unbalance, bent shaft, cracked shaft, misalignment, looseness, rubbing, gear faults, motor faults and impellor/blade defects. However, whilst these faults are common, they occur primarily in high speed machinery (> 600 rpm). There is limited information on faults which occur exclusively in low speed machinery, other than those initiated by rolling element bearing defects [3].

Numerous research works have been published on the detection and diagnosis of rolling element bearing defects. Tandon and Nakra [4] presented a detailed review of different vibration and acoustic methods, including vibration measurements in the time and frequency domains, sound measurements, the shock pulse method and the acoustic emission technique. Tandon and Choudhury [5] subsequently updated the reviews incorporating more recent works and advanced techniques currently being adopted in bearing defect detection. Most of the research on bearing diagnosis can be categorized in the time domain and frequency domains. The RMS, crest factor, probability density moments (skewness, kurtosis) are the most popular statistical time domain parameter for bearing defect detection [6-10]. In the frequency domain, the enveloping method, also known as demodulation or HFRT (High Frequency Resonance Technique), have been proven to be a very efficient and popular technique for detection of the characteristic frequencies of bearings [11-16]. Wavelet analysis has also been successfully applied to bearing defect detection [17-21]. In recent years, artificial neural networks and fuzzy logic have also emerged as popular tools for automated fault diagnosis [18, 22-25].

The ultrasonic technique has more recently been proposed as a powerful tool for condition monitoring of bearings in industrial applications particularly for low speeds. This technique captures and processes sound waves that manifest above the human hearing range and has been presdominantly applied in leak detection, crack detection, bearing defect detection and detecting lack of lubrication. Ultrasonic analysis is one of the less complex and less expensive CM techniques [26]. Its simplicity is directly related to the size and ease of use of handheld detectors, and the relatively straightforward presentation of measurement data on meters or digital readouts. However, a quantitative analysis of the signals from ultrasonic detector has not been conducted for condition monitoring of low speed bearings. Furthermore, traditional techniques involving vibration acceleration may not be able to detect an incipient failure due to the low impact energy generated by the relative moving components in low speed bearings.

In this study, the effectiveness of the ultrasonic technique for condition monitoring of low speed bearings is examined using on an experimental approach. An artificially induced fault in the form of a minute scratch on the bearing inner-race was utilised for this comparative study. The ultrasound signal from a normal bearing and a defective bearing are compared with vibration signals through time-domain statistical parameters and frequency-domain envelop spectrum at shaft speeds ranging from 1200 rpm to 30 rpm.

2 THE ULTRASOUND TECHNIQUE

Ultrasound is defined as sound waves that have frequency levels above 20 kHz; higher than what the unaided human ear can normally hear. Most machines emit consistent sound patterns under normal operating conditions. These sonic signatures can be defined and recognized; and changes in these signatures can be identified as components begin to wear or deteriorate. This enables technicians to identify and locate bearing deterioration, compressed air or hydraulic fluid leaks, vacuum leaks, steam trap leaks and tank leaks [26]. Airborne ultrasound operates in the lower ultrasonic spectrum of 20 kHz to 100 kHz. A compressed gas or fluid forced through a small opening creates turbulence with strong ultrasonic components on the downstream side of the opening. Also vacuum leaks produce turbulence similar to pressure leaks; however, the ultrasound is generated within the system. Poorly seated valves can also be detected.

It is worth noting that the Acoustic Emission (AE) technique also deals with signals in the high frequency range and has been increasingly used for condition monitoring of rotating machinery as well as structures. However, the AE technique differs from the ultrasonic technique in terms of the frequency range of interest and parameters for condition assessment. The AE technique generally operates in the 100 kHz to 1 MHz while the ultrasonic technique focuses on the frequency range of 20 kHz to 100 kHz. Parameters such as ring down counts, events, rising time, duration and peak amplitude are normally used in the AE technique to examine abnormality. In the ultrasonic detector, abnormality is usually detected by listening to the characteristics of sound or RMS indicator on the panel.

The heterodyne circuit is the main component in ultrasound detectors. It takes the ultrasound signal detected by the transducer, and converts it into an audible signal (< 20 kHz). This heterodyned signal can be enhanced by an audio amplifier to be heard using standard headphones, or can be processed through a converter to obtain a quantitative output in decibels (dB). The heterodyned signal can also be recorded through conventional data acquisition system.

This study uses the heterodyned structure-borne high frequency elastic waves between 20 - 100 kHz. The main advantage of ultrasound is the high signal-to-noise ratio detection ability. This advantage allows the exact localization of the energy source of the ultrasound activity, regardless of environmental interferences (i.e. noise) [27]. Although the ultrasonic technique using commercial ultrasound detectors is known to have the capability for detecting bearing defects, literature on the application of ultrasonic techniques for bearing condition monitoring using advanced signal processing techniques is generally absent. This study attempts to fill the gap where advanced signal processing techniques in the time and frequency domains are applied to signals obtained from both the ultrasonic technique and vibration measurements.

3 SIGNAL PROCESSING TECHNIQUES FOR BEARING CONDITION MONITORING

3.1 Time Domain Method

The time series signal can be used to perform fault and failure diagnosis by analysing vibration or acoustic data obtained from the equipment. Statistical methods are widely used to investigate the random characteristics of a physical system. It is important to be able to summarize the data obtained and be able to draw meaningful and useful results. The simplest method is to use overall root-mean-square (RMS) level and crest factor, i.e., the ratio of peak value to RMS. This method has been applied with limited success in the detection of localized defects [28-30]. Probability density has also been used popularly for bearing defect detection [8, 9, 29, 31, 32].

This study presents the analysis of vibration and ultrasonic signals with statistical parameters to detect the presence of defects in low speed bearings. The focus of this study lies in determining the relationship between the rotational speed of the bearing and the statistical parameters. This study presents the use of statistical parameters such as RMS, crest factor, skewness and kurtosis, to analyse an induced localized defect at low speeds.

- RMS –suited for steady-state signals
- Crest Factor – a trendable parameter which gives the ratio of the peak to RMS levels.
- Skewness – measures the relative energy above and below the mean level.

• Kurtosis –a compromise measurement between the insensitive lower moments and the over-sensitive higher moments. It is particularly useful in the detection of bearing failure.

3.2 Frequency Domain Method

The Frequency domain refers to the treatment of signals expressed as a function of frequency using the time domain signal. Whilst certain information is more easily interpreted in the time domain, the most detailed analysis of rotating machinery vibration data is often conducted in the frequency domain. Frequency-domain or spectral analysis of the vibration signal is perhaps the most widely used approach for bearing defect detection. It is often the case that signature spectral comparisons are not suitable for detecting damage to rolling element bearings. This is because the energy produced by the bearing defect is overpowered by more dominant signals from other rotating elements, nearby machinery and noise. For this reason, envelop detection is a commonly used signal processing technique for detecting incipient defects in rolling element bearings. This technique [32-34] involves a high-pass filtering operation to eliminate dominating low frequency components from the original signal. This signal is then rectified, demodulated, and a low-pass filter is finally used to eliminate the carrier high frequency. The processed signal is then displayed in the frequency spectrum showing the isolated bearing defect frequencies.

3.3 Modified Peak Ratio

In the detection of a localized bearing defect, it is essential to determine the presence and the severity of the defect. Enveloping technique or HFRT has proven to be an efficient method of displaying bearing defect frequencies by isolating other unwanted frequencies. However, for automatic diagnosis of bearing defect it is necessary to extract a symptom of failure without any human involvement. Shiroishi [32] suggested a peak ratio (PR) as an indicator to identifying the presence of bearing faults in the spectrum. PR is defined as the sum of the peak values of the defect frequency and harmonics over the average value of the spectrum and is shown in Eq. (1).

$$PR = \frac{N\sum_{j=1}^{n} P_j}{\sum_{k=1}^{N} S_k} \tag{1}$$

Where, P_j is the amplitude value of the peak located at the defect frequency harmonic, S_k is the amplitude at any frequency, N is the number of points in the spectrum, and n is the number of harmonics in the spectrum.

A *modified PR* is proposed in this study in order to make it more effective in showing the severity of the defects. The modified PR is defined in the following equation using the differences between the peak defect frequencies and the average value of the spectrum over the average of the spectrum.

$$mPR_O = 20\log_{10}\frac{\sum_{j=1}^{n}(P_j - A_s)}{A_s} \quad \text{(dB)} \tag{2}$$

$$mPR_I = 20\log_{10}\frac{\sum_{j=1}^{n}(P_j - A_s) + \sum_{i=1}^{l}(Ps_i - A_s)}{A_s} \quad \text{(dB)} \tag{3}$$

$$A_S = \frac{\sum_{k=a}^{b} S_k}{(b-a)}$$

Where, A_s is an average spectrum amplitude in the frequency band from a to b. By using a frequency band instead of whole spectrum band, PR can be a reliable indicator for earlier defect detection. In the case of an incipient bearing defect the amplitude of defect frequencies are often smaller than other peak frequencies.

4 EXPERIMENTS

The test rig used in this study is shown in Figure 1. The shaft is supported by two bearings and connected with an induction motor with a flange-bolt-rubber coupling. The shaft speed can be controlled by an inverter from 30 rpm to 6000 rpm. The bearings used in this test were cylindrical roller bearing (SKF, NJ2204) which has a removable inner ring. The localized defect was obtained by a minute scratch on the inner race. Previous studies have also involved seeded defects induced by using acid etching, spark erosion, scratching or mechanical indentation. Relatively small defects with a width of 1mm and average depth of 20 μm were used to assess the detection capability of incipient failures. It is well known that inner race defects are more difficult to detect than outer race defects due to the relative motion of the ring and the rolling elements. The bearing characteristic frequency for inner race fault (BPFI) is calculated to 6.713 times the shaft speed (Hz). The equipment involved in the collection of vibration and ultrasound data is shown in Fig. 1. It consists of an accelerometer (IMI 608A11), an ultrasound detector (UE Systems ultraprobe 9000), a signal conditioner (PCB 482A20), an analog filter (Krohn-Hite 3202), a DAQ Card (National Instruments 6062E) and a laptop computer with LABVIEW 7.1 (National Instruments).

Vibration and ultrasound data were collected at speeds ranging from 30 rpm to 1200 rpm. The data was analysed using statistical methods in the time domain and envelop spectra in the frequency domain. The vibration signal was sampled at 40 kHz with a cut-off frequency of 10 kHz. Ultrasound was sampled at 140 kHz with a cut-off frequency of 50 kHz. The time duration for each data set was 10 sec for the accelerometer and 4.5 sec for the ultrasound probe. The signal processing and analysis was achieved with the use of MATLAB 7.0 for feature extraction and enveloping.

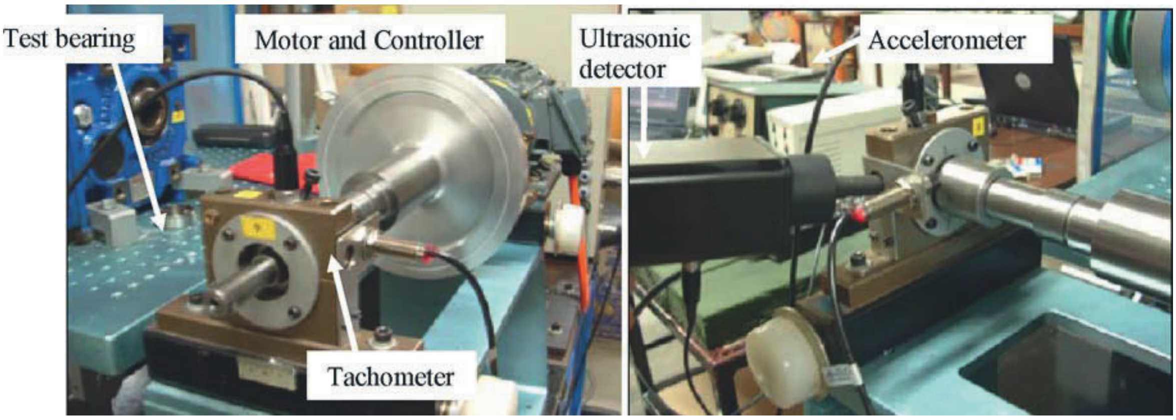


Figure 1. Experimental rig.

5 RESULTS

The signals measured in this experiment were vibration and the heterodyned ultrasound from a normal bearing and the bearing with the inner-race defect, bearing with a range of shaft speeds from 30 rpm to 1200 rpm. The objectives of this test and analysis were to gauge the effectiveness of the ultrasonic techniques by comparing it with vibration signals. The experimental tests are used to uncover the best statistical parameter for fault detection and to introduce an efficient frequency domain indicator for condition monitoring of low speed bearings.

5.1 Time wave form

Figure 2 shows the time wave forms of vibration acceleration signals from a normal and a defective bearing operating at low (150 rpm) and high speeds (1200 rpm). It was noted that at the low speed, there is no significant difference in the characteristics of signals from the normal bearing and from the defective bearing. The two signals can be easily distinguished at high speeds. This implies a potential limitation of the acceleration signal for fault detection of low speed bearing defects. Figure 3 shows ultrasound signal with a heterodyne frequency of 30 kHz under the same conditions. There are significant differences between a normal and a defective bearing in terms of magnitude and shape of the signal. Furthermore, the signals are much clearer at low speeds and display a number of impulses which are generated by the localized defects in bearings. By comparing these two figures, it can be concluded that ultrasound signals are more sensitive at detecting localized defects in bearings than acceleration signals.

5.2 Time Domain Parameters

In this paper, RMS, skewness, kurtosis and crest factor were used to compare their effectiveness for condition monitoring of low speed bearings on the basis that they are commonly used parameters for condition monitoring of bearings. For quantitative comparisons, all parameters are presented in decibel scale with reference of 1mV for each parameter. Figure 4 shows the comparisons of statistical parameters of vibration acceleration signals from a normal bearing with those from a defective bearing at different shaft speeds. RMS and skewness increased as the shaft speed decreased, while the kurtosis and crest factor decreased slightly. The relative amplitudes of parameters are shown in Figure 5. It is importantly noted that all the statistical parameters from vibration acceleration signals have smaller relative amplitudes as the shaft speed decreased, which indicates the vibration signal is not appropriate for condition monitoring of low speed bearings. The skewness and kurtosis values provided good detection at high speeds. Their detection abilities decreased sharply as speed decreased.

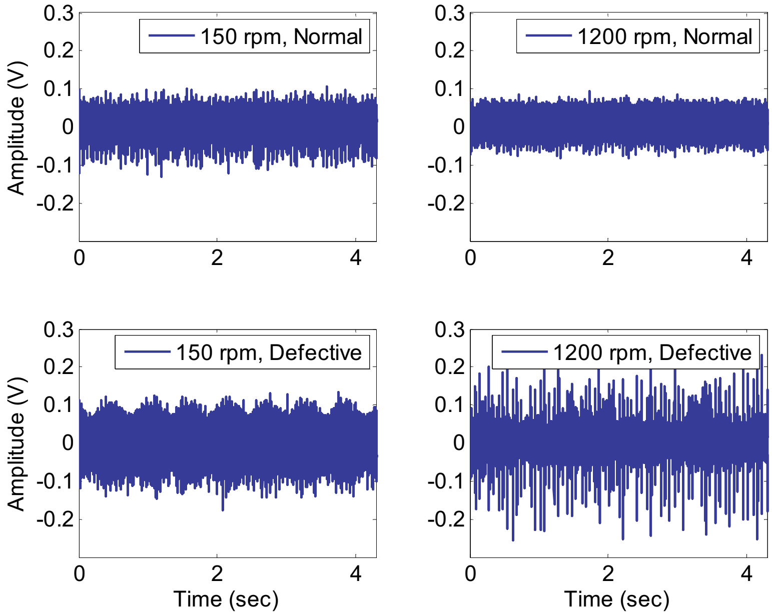


Figure 2. Time wave forms of acceleration signal from a normal and a defective bearing at 150 rpm and 1200 rpm.

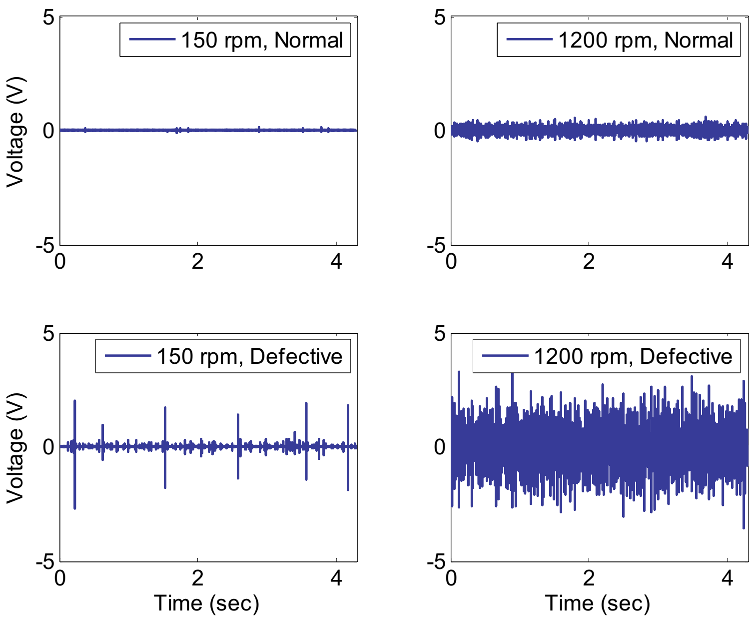


Figure 3. Time wave forms of ultrasound signals from a normal and a defective bearing at 150 rpm and 1200 rpm.

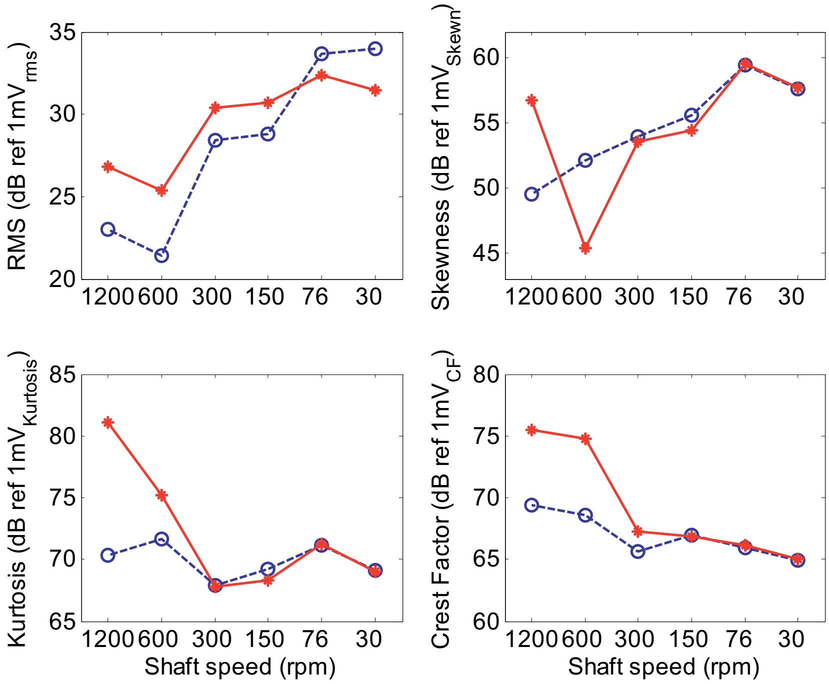


Figure 4. Comparison of statistical parameters from vibration (--o--:normal, —*—:defective).

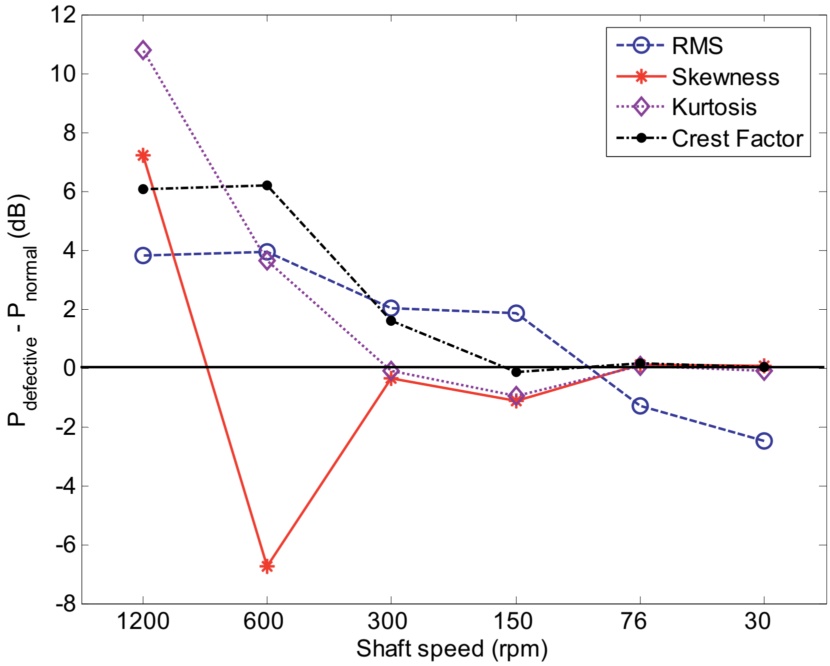


Figure 5. Relative amplitude of statistical parameters from vibration acceleration measurements.

Likewise comparisons of those parameters from ultrasound signals are presented in Figures 6 and 7. The RMS values almost linearly decreased according to a decrease in shaft speed with similar relative amplitudes at all speeds. This means that the RMS value is a good indicator for condition monitoring of bearings for the range of shaft speeds. This result reveals the reason why most ultrasound detector manufacturers use RMS as an indicator for condition monitoring. Skewness didn't show a clear trend for the range of speeds. In contrast, crest factor and kurtosis showed interesting results. Their values from a defective bearing were smaller than those from a normal bearing at high speeds. However, when shaft speeds were less than 300 rpm their relative values increased dramatically. This clearly indicates its significant potential for condition monitoring of slow speed bearings.

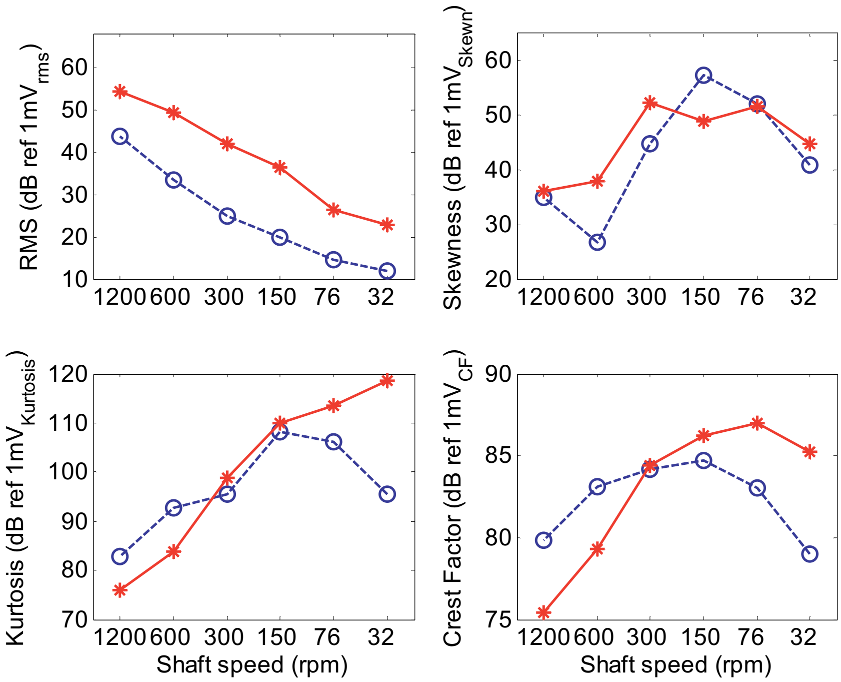


Figure 6. Comparison of statistical parameters from the ultrasound signal (--o--:normal, —*—:defective).

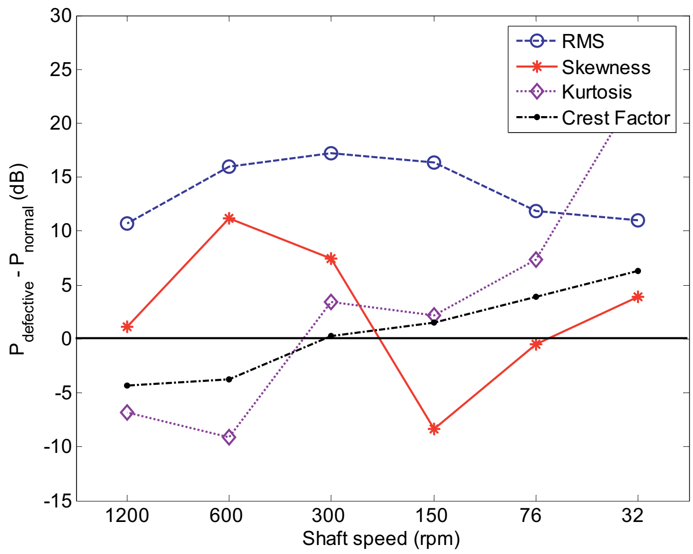
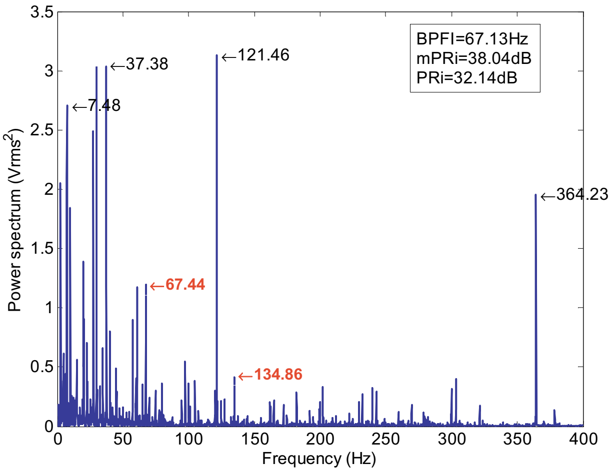


Figure 7. Comparison of differences of statistical parameters from the ultrasound signal.

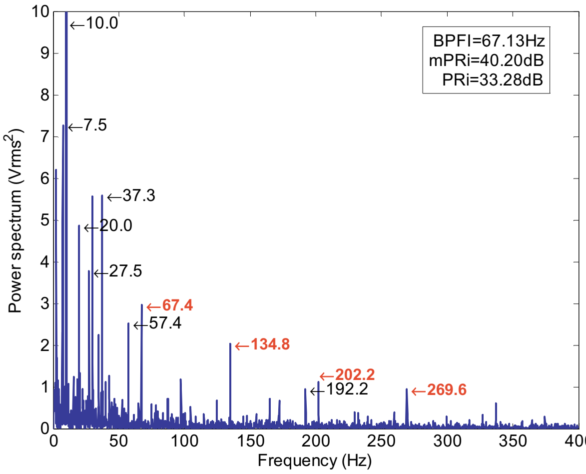
It is worth noting that RMS of the ultrasound signals is a very reliable indicator at most speeds showing relatively constant amplitudes larger than 10dB. The kurtosis value of ultrasound signals is potentially the best parameter at very slow speeds (less than 50 rpm), followed by crest factor. By comparing the results from ultrasound signals in Figure 7 with those from acceleration signals in Figure 5, it is evident that the ultrasound technique is superior to vibration measurements at low speeds.

5.3 Frequency Domain Parameters

As explained in the previous section, the modified Peak Ratio is introduced in this paper to improve the reliability of identifying the presence of bearing defect characteristic frequencies in the frequency spectrum. Figure 8 shows the enveloped power spectrums of vibration acceleration signals and ultrasound signals at 600 rpm of a defective bearing. The defect frequency (BPFI=67.13 Hz) and its harmonics are clearly evident in both spectra. However, they are more significant in the ultrasound spectrum with relevant side bands of the shaft speed. It is also noted that the modified Peak Ratio for the inner-race fault frequency (*mPRi*) which is proposed in this study, has a much higher value (6-7dB) than the original Peak Ratio proposed by Shiroishi [32], implying that *mPRi* can be an effective indicator for the detection of bearing defect frequencies in frequency spectra.



(a) vibration



(b) ultrasound

Figure 8. Enveloped power spectrum from defective bearing at 600 rpm.

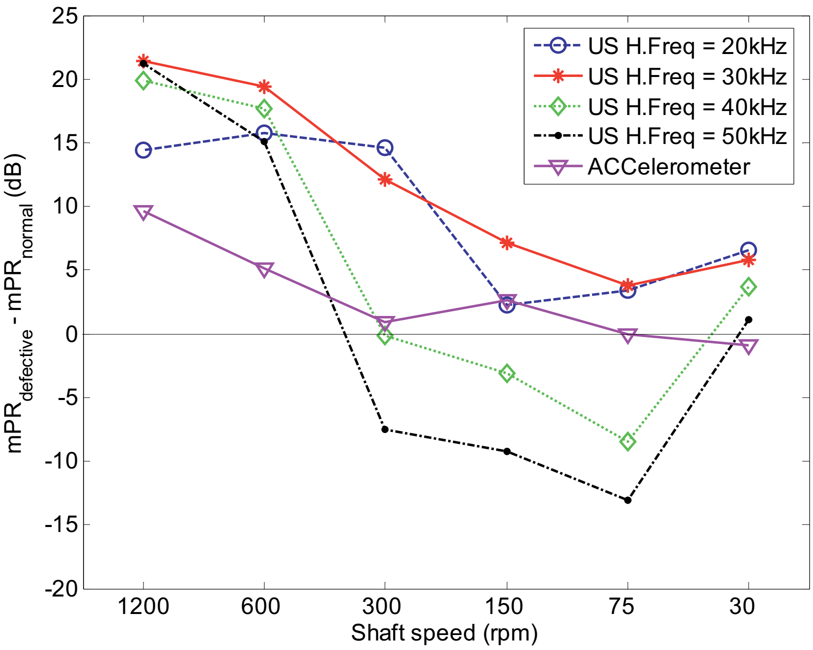


Figure 9. Relative modified Peak Ratio from ultrasound and acceleration.

Finally, the effectiveness of the *mPR* from ultrasound signals was compared with those from vibration signals and the results are shown in Fig. 9. In this study four different heterodyne frequencies in the ultrasound detector were examined for comparison. The 30 kHz heterodyne frequency gives the best result at most speeds and followed by 20 kHz. It is also clear from this study that the vibration acceleration measurements are inferior to the ultrasound technique in terms of detectability of fault in the frequency domain at all shaft speeds.

6 CONCLUSIONS

A number of experiments have been carried out to investigate the effectiveness of the ultrasonic techniques for condition monitoring of low speed bearings. A defective bearing which a simulated defect on the inner race was used in conjunction with a healthy bearing at different shaft speeds ranging from 1200 rpm to 30 rpm. The vibration signals obtained from an accelerometer were also measured and analysed for comparative purposes. The time-domain statistical parameters and frequency-domain modified Peak Ratio were calculated and compared. This study revealed that ultrasound technique is demonstrably superior to vibration acceleration measurements for detecting incipient defects in low speed bearings. The RMS of ultrasound signals provided the best parameter at almost all speeds. However, at very low speeds (less than 50 rpm), the kurtosis and crest factors performed best. In the frequency domain, a modified Peak Ratio was proposed and was proven to a better indicator than the original Peak Ratio. A setting of 30 kHz for the heterodyne frequency on the ultrasound detector gave the best result. Again, detection of fault from ultrasound signals was superior when comparisons were made in the frequency domain. The work is progressing in getting more samples from both the laboratory and the field and these results will be presented in due course.

7 REFERENCES

[1] R. James, "Machine Condition Monitoring...A Catalist for Change in Industrial Equipment Maintenance," *P/PM Technology*, vol. October, pp. 20-40, 2001.

[2] R. James, "Machine Condition Monitoring...A Catalist for Change in Industrial Equipment Maintenance," *P/PM Technology*, vol. December, pp. 18-37, 2001.

[3] J. E. Berry, *Required Vibration Analysis Techniques and Instrumentation on Low Speed Machines*: Technical Associates of Charlotte, PC, 1997.

[4] N. Tandon and B. C. Nakra, "Vibration and acoustic monitoring techniques for the detection of defects in rolling element bearings - a review," *Shock and Vibration Digest*, vol. 24, pp. 3-11, 1992.

[5] N. Tandon and A. Choudhury, "A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings," *Tribology International*, vol. 32, pp. 469-480, 1999.

[6] S. Braun and B. Datner, "Analysis of Roller/Ball Bearing Vibrations," *Transactions of the ASME*, vol. 101, pp. 118-125, 1979.

[7] R. J. Alfredson and J. Mathew, "Time Domain Methods for Monitoring the Condition of Rolling Element Bearings," presented at Mechanical Engineering Transactions, 1985.

[8] D. Dyer and R. M. Stewart, "Detection of Rolling Element Bearing Damage by Statistical Vibration Analysis," *Journal of Mechanical Design, Transaction of the ASME*, vol. 100, pp. 229-235, 1978.

[9] J. Mathew and R. J. Alfredson, "The Condition Monitoring of Rolling Element Bearings Using Vibration Analysis," *Journal of Vibration, Acoustics, Stress, and Reliability in Design, Transaction of ASME*, vol. 106, 1984.

[10] N. Tandon and B. C. Nakra, "Detection of Defects in Rolling Element Bearings by Vibration Monitoring," *IE(I) Journal MC*, vol. 73, pp. 271-282, 1993.

[11] T. I. Liu and J. M. Mengel, "Intelligent Monitoring of Ball Bearing Conditions," *Mechanical Systems and Signal Processing*, vol. 6, pp. 419-431, 1992.

[12] R. J. Alfredson and J. Mathew, "Frequency Domain Methods for Monitoring the Condition of Rolling Element Bearings," presented at Mechanical Engineering Transactions, 1985.

[13] J. I. Taylor, "Identification of Bearing Defects by Spectral Analysis," *Journal of Mechanical Design, Transaction of the ASME*, vol. 102, pp. 199-204, 1980.

[14] P. D. McFadden and J. D. Smith, "Vibration monitoring of rolling element bearings by the high frequency resnance technique - a review," *Tribology International*, vol. 17, pp. 3-10, 1984.

[15] Y. A. AZOVTSEV, A. V. BARKOV, and L. A. YUDIN, "Automatic Diagnostics and Condition Prediction of Rolling Element Bearings Using Enveloping Methods," presented at Vibration Institute Proceedings, 18th Annual Meeting, Hershey, Pennsylvania, June 21-23., 1994.

[16] A. V. Barkov, N. A. Barkova, J. S. Mitchell, and S. J. Capistrano, "Assessing the Condition and Lifetime of Rolling Element Bearings from a Single Measurement," presented at Vibration Institute Proceedings, 19th Annual Meeting, Indianapolis, Indiana, June 20-22, 1995.

[17] J. Lin, M. J. Zuo, and K. R. Fyfe, "Mechanical Fault detection Based on the Wavelet De-Noising Technique," *Journal of Vibration and Acoustics, Transactions of the ASME*, vol. 126, pp. 9-16, 2004.

[18] B. A. Paya, I. I. Esat, and M. N. M. Badi, "ARTIFICIAL NEURAL NETWORK BASED FAULT DIAGNOSTICS OF ROTATING MACHINERY USING WAVELET TRANSFORMS AS A PREPROCESSOR," *Mechanical Systems and Signal Processing*, vol. 11, pp. 751-765, 1997.

[19] J. LIN and L. QU, "FEATURE EXTRACTION BASED ON MORLET WAVELET AND ITS APPLICATION FOR MECHANICAL FAULT DIAGNOSIS," *Journal of Sound and Vibration*, vol. 234, pp. 135-148, 2000.

[20] P. W. Tse, Y. H. Peng, and R. Yam, "Wavelet Analysis and Envelope Detection For Rolling Element Bearing Fault Diagnosis - Their Effectiveness and Flexibilities," *Journal of Vibration and Acoustics, Transactions of the ASME*, vol. 123, pp. 303-310, 2001.

[21] B. Liu, S. F. Ling, and R. Gribonval, "Bearing failure detection using matching pursuit," *NDT & E International*, vol. 35, pp. 255-262, 2002.

[22] M. Subrahmanyam and C. Sujatha, "Using neural networks for the diagnosis of localized defects in ball bearings," *Tribology International*, vol. 30, pp. 739-752, 1997.

[23] B. Li, G. Goddu, and M.-Y. Chow, "Detection of common motor bearing faults using frequency-domain vibration signals and a neural network based approach," presented at American Control Conference, 1998. Proceedings of the 1998, 1998.
[24] C. K. Mechefske, "OBJECTIVE MACHINERY FAULT DIAGNOSIS USING FUZZY LOGIC," *Mechanical Systems and Signal Processing*, vol. 12, pp. 855-862, 1998.
[25] B. Li, M.-Y. Chow, Y. Tipsuwan, and J. C. Hung, "Neural-network-based motor rolling bearing fault diagnosis," *Industrial Electronics, IEEE Transactions on*, vol. 47, pp. 1060-1069, 2000.
[26] "Ultraprobe Introduction Manual, available : www.uesystems.com," UE systems.
[27] T. J. Holroyd, *The Acoustic Emission & Ultrasonic Monitoring Handbook*. Oxford: Coxmoor Publishing Company, 2000.
[28] F. Bolaers, O. Cousinard, P. Marconnet, and L. Rasolofondraibe, "Advanced detection of rolling bearing spalling from de-noising vibratory signals," *Control Engineering Practice*, vol. 12, pp. 181-190, 2004.
[29] R. B. W. Heng and M. J. M. Nor, "Statistical Analysis of Sound and Vibration Signlas for Monitoring Rolling Element Bearing condition," *Applied Acoustics*, vol. 53, pp. 211-226, 1998.
[30] J. Shiroishi, Y. Li, S. Liang, S. Danyluk, and T. Kurfess, "Vibration Analysis for Bearing Outer Race Condition Diagnostics," presented at Journal of Brazilian Society of Mechanical Science, Rio de Janeiro, 1999.
[31] N. Tandon, G. S. Yadava, and K. M. Ramakrishna, "A comparison of some condition monitoring techniques for the detection of defect in induction motor ball bearings," *Mechanical Systems and Signal Processing*, vol. In Press, Corrected Proof.
[32] J. Shiroishi, Y. Li, S. Liang, T. Kurfess, and S. Danyluk, "Bearing Condition Diagnostics via Vibration and Acoustic Emission Measurements," *Mechanical Systems and Signal Processing*, vol. 11, pp. 693-705, 1997.
[33] P. D. McFadden and J. D. Smith, "Vibration monitoring of rolling element bearings by the high frequency resonance technique - a review," *Tribology International*, vol. 17, pp. 3-10, 1984.
[34] N. Tandon and B. C. Nakra, "Comparison of vibration and acoustic measurement techniques for the condition monitoring of rolling element bearings," *Tribology International*, vol. 25, pp. 205-212, 1992.

8 ACKNOWLEDGMENTS

The authors wish to acknowledge UE System’s sponsorship of the Ultraprobe 9000. The project is supported by the CRC for Integrated Engineering Asset Management (CIEAM).

DEVELOPMENT OF TESTING FACILITIES FOR VERIFICATION OF MACHINE CONDITION MONITORING METHODS FOR LOW SPEED MACHINERY

V Kosse & A Tan

School of Engineering Systems, Queensland University of Technology
2 George St., GPO Box 2434, Brisbane, Qld 4001, Australia

Abstract: Machine condition monitoring (MCM) of low speed machinery is highly challenging task. Methods of MCM of machinery operating at a speed of 500 to 3000 rpm are well-developed including wear debris and vibration monitoring (both instrumentation and signal processing). Machinery operating at a speed below 500 rpm is usually grease lubricated and wear debris analysis cannot be effectively used. Monitoring of vibrations at such speed is very difficult because conventional accelerometers give very week signal especially at a speed below 100 rpm. There is a vital need to develop new approaches to monitoring fault development on low speed machinery, which include new sensors and alternative methods such as sound emission and ultrasonic emission. Essential part of this process is development of testing facilities that enable modeling of different combinations of loading conditions that take place on low speed machinery and verify new MCM methods. For this purpose a design study has been conducted to develop methods of modeling of different combinations of loads on bearings and gears operating at low speed. This study resulted in the development of a unique highly versatile test rig presented in this paper. It enables modeling of different types of bearings operating at a speed of 30 to 600 rpm under a combination of different loading conditions, in particular, steady load, impact load, swinging load, axial load and rumble. It also enables modeling of transitional processes in shaft-mounted, flange-mounted and foot-mounted gear drives under different loading conditions (start up, coast down, increase or decrease of speed). Special instrumentation enables monitoring of the supply frequency, phase current, power consumed and taking these parameters for recording.

Key Words: Machine condition monitoring

1 INTRODUCTION

Low speed machinery is widely used in industry. The speed is considered low if the machine operates at a speed below 600 r.p.m. Some examples are: metallurgical converters and mixers, cement furnaces, kilns, crushers, etc. To drive such machines multi-stage gear drives are used with DC or frequency-controlled AC motors. Such machines quite often experience shock loading, frequent star and stop. Open gear sets may entrap lamps of hard dirt and crushing it induces rumble (vibrations) in the drive train. The dynamic factor on such machinery quite often exceeds two especially during transitional processes. An accidental failure on such machinery may have disastrous consequences both in terms of down time, lost production and environmental impact. Large safety factors and over designing the machinery helps to a certain degree but contributes to increased cost of equipment.

Machine Condition Monitoring (MCM) methods for high speed machinery are well developed. They include vibration monitoring, wear-debris analysis, thermography, sound emission and ultrasonic diagnostic methods. MCM methods for low speed machinery remain “undigged field” as a result of challenges discussed in the next section. Development of MCM methods for low speed machinery is one “burning” issue. Another problem is that at this point in time there are no experimental facilities that enable investigation of different mechanical faults development at low speed and verification of MCM methods for such faults.

In this paper approaches to experimental investigation of different mechanical faults development at low speed are discussed and a concept of a low speed test rig is presented.

2 CHALLENGERS TO MACHINE CONDITION MONITORING OF LOW SPEED MACHINERY

Mechanical drives for low speed machinery are usually three or four stage gearboxes. If a lubricant is selected for an intermediate stage, high speed input stage and low speed output stage of the gearbox will be inadequately lubricated. In long term it makes the gearbox vulnerable to excessive wear and gear fault development. Gear drives for mixers processing powdered products quite often have a reciprocating mechanism built in the gearbox, which causes shock loading. Many intermediate shafts have bearing supports inaccessible for sensors and suppliers of drives would not allow to drill holes in the gearbox to take cables out.

Drives on mixers and crushers are typically shaft mounted or cradle mounted and there is a torque arm that holds the reactive torque. During transition mode (start up, coast down or load variations) knocking signal is generated in the torque arm joints, which is superimposed on harmonics generated by the machine itself.

If vibrations are used for MCM purpose, energy of signals coming from high speed stages and shock loading of the operating unit is high, and it masks low energy signal coming from low speed stages. Background noise makes situation even worse. There is a theoretical possibility to extract a low energy useful signal using signal processing methods, but in reality it is problematic.

The use of wear debris analysis is very difficult because low speed units usually are geese lubricated. In multi-stage gearboxes inadequate lubrication causes uneven wear, which is difficult to predict and monitor.

The use of thermography is problematic because the heat energy dissipation rate is comparable with the heat generation rate in the area of fault development and it is very difficult to detect notable temperature rise.

Main components of low speed drives that require attention from the standpoint of MCM are bearings and gears. Occasionally other components fail (e.g. shafts or mixing screws). Rolling elements bearings on low speed heavy machinery usually have cylindrical or taped rollers as rolling elements. Often double-raw rolling bearings are used, which makes the vibrational signature even more complex. With modelling of bearing faults in laboratory environment contradictory requirements have to be satisfied. In particular, real bearings on heavy machinery are quite large (300mm and bigger); on the contrary, to keep the cost of testing facilities down smaller bearings should be used. In general, to transfer results from a model to a real scale object “cauchi condition” can be used to determine multiplier coefficients for different parameters to satisfy conditions of dynamic modelling. However, fidelity of the modelling remains questionable if too small bearings are used on a model.

In industrial environment different combinations of loading on bearings may take place:

- Steady radial and / or axial load when the load does not vary in time.
- Impact (shock) loading.
- Cyclic loading.
- Rumble when radial or torsional vibrations are imposed on the shaft and transferred to the bearings.

When low speed bearing endurance test is conducted, the test rig has to be run continuously for days or even weeks. Maintaining the consistence of the load becomes a highly challenging task. In the case of a steady load the use of hydraulic devices is questionable because they are expensive, not suitable for long continuous use, and finally, it is almost impossible to maintain the value of the load within a narrow range (say ±2%). The use of weights is possible if the load required is measured in dozens or hundreds of Newtons. The use of larger weights becomes highly unpractical and even dangerous.

Impact or cyclic loading can be applied to the shaft using hammering mechanism. The challenge is how to make sure that the specific bearing is subjected to the shock load but not all bearing on the shaft of the test rig.

For modelling of rumble loading there are no any suggestions in research publications. The use of “saw tooth” disks with spring loaded ratchets is possible, but a challenge is how to control shock impulses in the radial and tangential directions.

On heavy machinery bearing rings may rotate continuously, may perform swinging motion, or one of the rings may remain stationary. The challenge is how to reproduce in the laboratory environment all these kinds of motion under a combination of steady and shock loading.

A real challenge is how to initiate a controlled fault development. There are three possibilities:

- Bearing overloading. This approach is relatively easy to implement, but difficult to pick the moment of the fault development.
- Contaminated lubricant. This method has been suggested by B.Kuhnell [1, 2]. It is easy to implement, but difficult to control.
- Introduced fault. The most common method is the spark erosion. It enables to make a notch of a known size, which is approximately about 0.5mm in width. It is however, too large compared to real faults developing on bearing race surface.

Simulation of gear faults development in laboratory environment requires the satisfaction of the following conditions:

- Provision for shaft-, flange- and foot-mounted gearboxes.
- Possibility of visual monitoring of the fault development.
- Possibility to reconfigure the test rig from gearbox test to bearing test.
- Accessibility of housing surfaces for sensor mounting.

These requirements could be relatively easy satisfied by proper selection of gearboxes. The main challenge is how to load the gearbox with a measured torque. If a single gearbox is used with large reduction ratio (e.g. 100:1) the output torque will be 100 times the input torque. To apply that large torque magnetic particle brakes or eddy current brakes can be used. The magnetic particle brakes can maintain the loading torque for a long period of times (days is necessary). The main limitation is very high cost. The eddy current brakes are not suitable for prolonged use. As a consequence of heat released in the brake the loading torque decreases by 20-25%.

A solution to overcome this problem was suggested by V.Kosse [3]. Two gearboxes can be connected back-to-back with low speed shafts. In this case both the input and the output shafts rotate at high speed and relatively small torque is required to load the drive train. The following methods can be used to apply a loading torque at high speed:

Connect the drive train to a DC motor / generator. It is relatively simple. The main challenges are how to control the loading torque and how to dissipate significant amount of heat generated.

Connecting the drive train to a water pump (e.g. centrifugal pump). In this case very difficult to control the loading torque, a large water tank and the pipeline are required.

Disk brakes. Such brakes are widely used on large four-wheel drive vehicles (e.g. Toyota Land Cruiser). The disks are placed in a housing filled with oil. Such brakes can be used only for a short period of time, and the loading torque is difficult to control.

Rheological couplings. They are filled with a rheological fluid that offers different resistance to a disk rotation when the strength of a magnetic field generated by an electric magnet is changing. The strength of the magnetic field and the loading torque are easy to control and theoretically they can be used for a long period of time, but practically the rheological fluid quickly degrades (especially when heat is generated) and has to be frequently replaced, which on most rheological coupling is difficult to implement.

The technological problems discussed in this section show that the development of testing facilities for low speed machinery is a real scientific and technological challenge. In the next section possible approaches to simulation of the fault development in bearings and gear drives operating at low speed will be discussed.

3 SIMULATION OF DIFFERENT MECHANICAL FAULTS ON LOW SPEED MACHINERY IN LABORATORY ENVIRONMENT

Spectra Quest Incorporation (USA) developed a demonstrational rig for medium to high speeds. It is used manly for educational purpose and enables demonstration of most mechanical faults. It is table-mounted and has an AC frequency controlled motor driving through a coupling a shaft with two disks. The shaft rests on two bearing supports that can be shifted in the lateral direction. A small reciprocating mechanism can be attached to the shaft through a belt dive. The rig enables demonstration of the following faults: imbalance, shaft and coupling misalignment, rolling bearings faults, “cocked” rotor, resonance and critical speed study, sleeve bearing study, belt drive performance studies, mechanical rub, gearbox faults, reciprocating mechanism study, foundation effect study. The Spectra Quest demo rig is depicted in Figure 1. Attempts to use it at low speed are unsuccessful because the frequency control does not allow to maintain speed less than 100 r.p.m. Then, components of the rig are so tiny that make it impossible to apply loading conditions typical for low speed machinery.

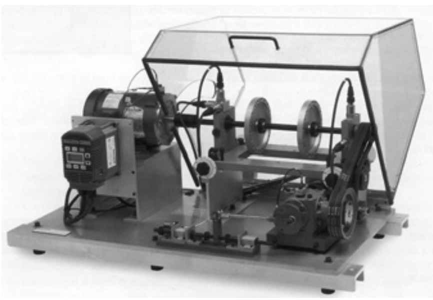


Fig. 1. The machinery fault simulator from Spectra Quest Incorporation (USA)

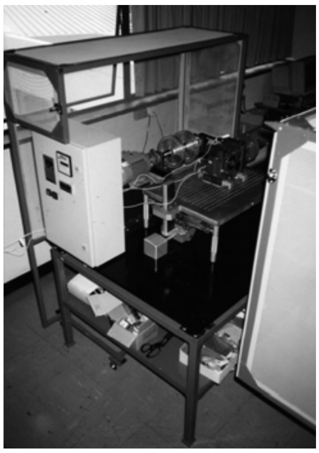


Fig. 2. QUT Fig. 1. The machinery fault simulator from Spectra Quest Incorporation (USA).

At Queensland University of Technology (QUT) V.Kosse developed a much bigger machinery fault simulator (MFS) [3] shown in Figure 2, that utilizes similar concept. The main difference is that it has a tilt table that can be moved parallel to the shaft and tilted in the lateral direction. This enables installation of different gearboxes and loading devices and simulation of belt misalignment faults. As a loading device a disk brake from Toyota Land Cruise is used. This MFS enables modeling of all typical mechanical faults at a speed from 60 to 1400 r.p.m. However, it cannot be effectively used for modeling of mechanical faults at low speed because the frequency controlled AC motor at low speed has such a small power output that any load applied to the shaft stops it. This proves that for simulation of mechanical faults at low speed a specialized test rig has to be developed.

3.1 Simulation of bearing faults

The most feasible concept for simulation of bearing faults at low speed is to use a shaft on two supports. At the driving end the shaft is attached to a gear motor through a coupling that can be quickly disconnected (e.g. chain coupling). The bearing support at the driving end has to be over-designed to ensure it will not fail under the test loading. The bearing support at the opposite end of the shaft (driven end) has to have a provision for installation of a single bearing or two bearings (e.g. two taper roller bearings in X-pattern). These bearings can be of different kinds but with the same bore and outer diameters.

The radial load can be applied close to the driven end support where the test bearings are installed through a rod preloaded with Belleville springs. The value of the loading radial force can be measured by a load cell. This will enable to maintain a constant loading force for a long period of time (days and weeks if necessary). The axial loading force can be applied if two taper roller bearings are installed in X-pattern through Belleville springs preloaded with the nuts. The value of the axial loading force can be monitored through the deflection of Belleville springs. Impact loading can be applied to the shaft through the hammering pivoted arm loaded with weights and cyclically moved with a cam mechanism.

Continuous rotation can be provided through a frequency-controlled gear motor with large reduction ratio. This will enable to maintain a low speed of the shaft rotation within the range of 30 to 600 r.p.m. without sacrificing the power output. Speed can be measured by means of a tachometer. Swinging motion of the shaft can be maintained by a crank mechanism driven by a small electric motor. The value of the swing angle can be controlled by varying the length of the crank arm.

The shaft assembly with the drive and loading devices can be mounted on a robust frame.

For the purpose of gearbox fault development monitoring at the driven end of the shaft another gearbox can be installed. This gearbox has to have a hollow low speed shaft and a flange. This will enable mounting of this gearbox with different configurations when the driving shaft is plugged in the low speed shaft of the gearbox:

- Flange (foot)-mounted when the flange is bolted to the frame.
- Shaft-mounted with the torque arm attached to the frame.

High speed shaft of this gearbox can be attached to a loading device. The gearbox will step up the speed and enable the loading of the whole drive train with a relatively low torque.

Analysis of possible methods of loading conducted above shows that the main problem is how to maintain the torque if experiment has to be conducted for several days of weeks continuously. Most loading devices are suitable for use only during sort period of time. At QUT a new loading device has been developed that utilizes a gear pump attached to the driven end of the transmission that pumps oil through a hydraulic station (see Figure 3). The loading torque can be easily adjusted by means of a throttling device. Large amount of oil (60 liters) enables the use of the loading device for at least of half an hour without a danger of overheating the oil. If the loading device has to be used for a long period of time, the heat exchanges can be used with flowing water that maintain the oil temperature within required range. The pump capacity is selected to match the power output of the driving gearbox and enables the overloading of the driven gearbox if necessary. The power consumed can be measures by means of a multi-functional meter attached to the AC motor, which also enables measurement of the phase voltage and current, the supply frequency and these parameters can also be taken to the recording device.

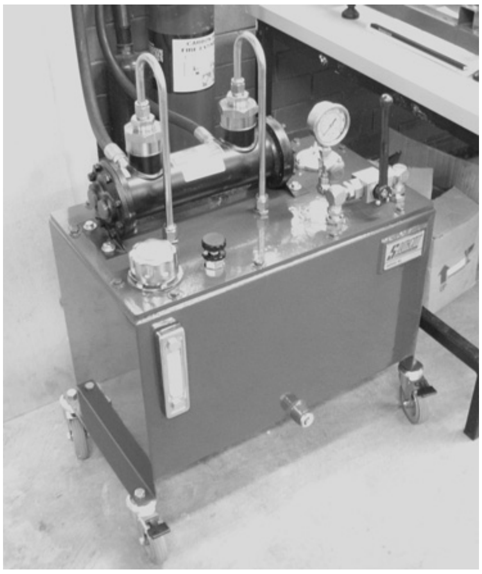


Fig. 3. Hydraulic station of the loading device (heat exchanger is seen on the top).

The approaches for simulation of different bearing and gear faults discussed above can be utilized on a test rig for low speed machinery. Such a test rig has been developed at QUT and is presented in the next section.

4 THE CONCEPT OF THE TEST RIG

A unique test rig for low speed machinery developed at QUT is shown in Figure 4. It has a robust frame on which major components and assemblies are mounted. The control box is not shown to enable clear view of all systems and components. The drive and the shaft assembly on two bearing supports are at the right hand side of the picture.

The hammering mechanism for modelling of impact loads is seen at the top. The rumble mechanism is represented with a profiled disk and spring-loaded roller. The shaft-mounted gearbox at the driven end is shown without loading device. The availability of a driving gearbox with large reduction ratio and frequency control enables to maintain steady power output at any speed from a fraction of revolution per second to 600 r.p.m. This test rig will enable simulation of typical mechanical faults at low speed and verification of different condition monitoring methods.

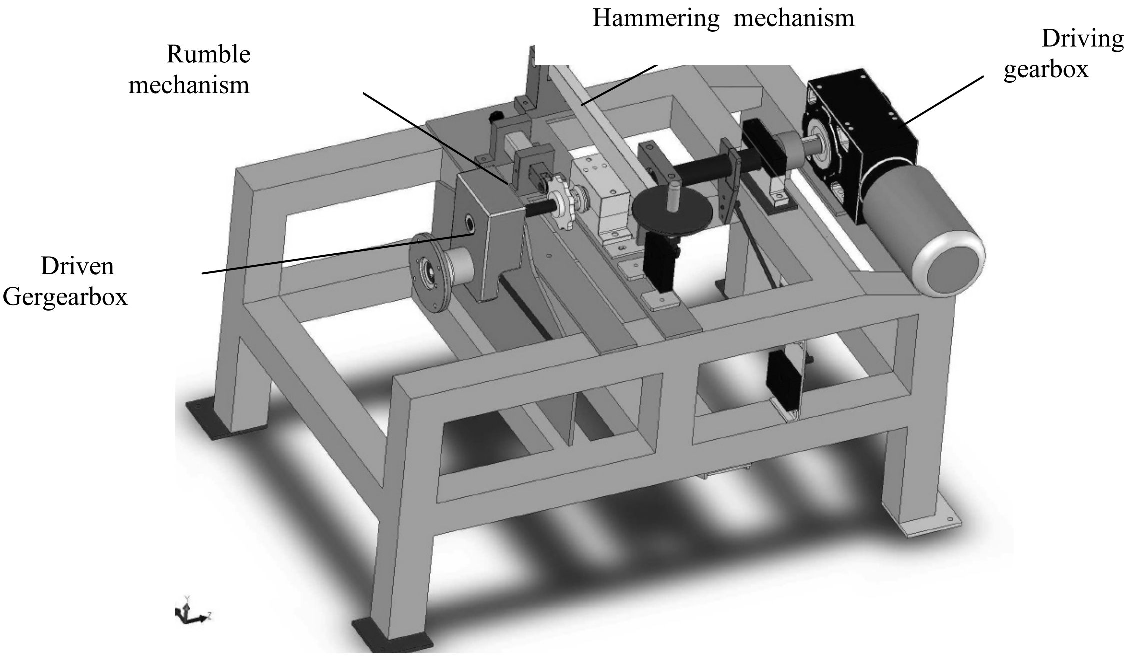


Fig. 4. Test rig for physical simulation of mechanical faults at low speed (30 to 600 r.p.m.).

5 CONCLUSIONS

In this paper approaches to experimental investigation of typical mechanical fault development have been discussed.

It is shown that simulation of typical loading conditions present on low speed machinery is highly challenging task both from scientific and technological standpoints.

Conceptual approaches to physical simulation of bearing and gear faults development on low speed machinery have been discussed.

A unique test rig developed at QUT for physical simulation of fault development on low speed machinery is presented.

This test rig will enable a systematic research on the development of machine condition monitoring methods for early fault detection on low speed machinery.

6 ACKNOWLEDGEMENT

The authors would like to acknowledge that the development of the testing facilities for low speed machinery has become possible with the support of the Cooperative Research Centre for Integrated Engineering Asset Maintenance (CIEAM) and Comalco Australia Ltd. through a project ID202.

7 REFERENCES

[1] Kuhnell, B.T. "Effects of age and contamination on rolling bearings and gears: Part 1". Proceedings of the 3-rd Asia-Pacific Conference on Systems Integrity and Maintenance (ACSIM2002), September 25-27, 2002, Hilton, Cairns, Australia.

[2] Kuhnell, B.T. “Effects of age and contamination on rolling bearings and gears: Part 2”. Proceedings of the 3-rd Asia-Pacific Conference on Systems Integrity and Maintenance (ACSIM2002), September 25-27, 2002, Hilton, Cairns, Australia.

[3] Kosse, V. and Tan, A. “Development of Experimental Facilities for Modelling of Typical Mechanical Faults”. Proceedings of the 3-rd Asia-Pacific Conference on Systems Integrity and Maintenance (ACSIM2002), September 25-27, 2002, Hilton, Cairns, Australia.

BLIND DECONVOLUTION USING THE EIGENVECTOR A LGORITHM (EVA) METHOD FOR THE ENHANCEMENT OF BEARING SIGNAL THROUGH THE TRANSMISSION CHANNEL

Andy C C Tan[a], M Karimi[a], J Mathew[b]

[a]*School of Engineering Systems, Queensland University of Technology, George Street, Brisbane Q4001, Australia*

[b]*CRC for Integrated Engineering Asset Management, Queensland University of Technology, QLD 4001, Australia.*

Abstract: Detection of an incipient internal defect from a measurement point on the surface of a machine is often made difficult due to the corruption of the source signal by background noise usually through the transmission path. In order to detect or recover the source signal, it is essential to suppress or remove this background noise. With the advancement of digital signal processing techniques, this removal can be achieved using fixed or adaptive filters. The main problem with the application of filtering techniques is that one has to know the characteristics of the noise in advance or may require more than two channels of data if using an adaptive filter. This paper reports on the application of Blind Deconvolution method to recover fault information from the measured (observed) signals of a damaged bearing through the transmission channel and which was also corrupted by noise.. The Blind Deconvolution method manages to recover the fault signal by estimating the inverse of the transmission (channel transfer function) path using the eigenvector algorithm (EVA). Modified crest factor (MCF) and Arithmetic Mean (AM) algorithms were used to optimise the equalizer parameters. Computer simulation and experimental studies were used to verify the applicability of the technique.

Key Words: Blind Deconvolution, Bearings, Condition Monitoring, Signal Processing, Vibration

1 INTRODUCTION

Rolling element bearings from one of the major components in rotating machinery and facilitates rotating motion by usually carrying heavy load over a small surface area. Although the life of rolling element bearings can be estimated based on the load and the rotating speed, catastrophic failure of these bearings can occur prematurely due to unpredictable loading or duty or even vibration transmitted from other machine components. Other factors which can influence the life of a bearing are lubricant film, static and dynamic load, geometry of the housing, shaft and bearing, pre-load and maintenance personnel [1]. One alternative to preventing catastrophic failure is to replace the bearings after a certain period of operation through planned maintenance.

Numerous techniques are now available to detect rolling element bearing failures but the suitability of the techniques varies from system to system. Detection of internal malfunction via vibration analysis is widely used. Some of the techniques commonly used in condition monitoring of rolling element bearings include shock pulse, statistical analysis (crest factor, kurtosis and RMS level) and frequency based detection techniques [2-4]. Unfortunately, vibration analysis based on frequency and time domain analysis can only be effectively applied when the signal-to-noise ratio is relatively high. Although these techniques are widely used, they are now mainly used at the final stage of the analysis process after an improvement in the signal-to-noise ratio is achieved.

In situations where the vibration signal is corrupted by background vibration (referred to as noise) conventional analysis techniques may fail to detect an incipient failure. At this stage the bearing signal is generally small and its detection would require the noise component in the overall vibration signal to be severely attenuated with a consequential improvement to the signal to noise ratio. One of the earliest methods to suppress noise is the application of adaptive filtering techniques, such as adaptive noise cancellation (ANC)[5,6]. This technique is also used in a number of applications, namely, the suppression of unwanted noise and detection of tool wear acoustic emission signals. Other applications of adaptive signal processing include identifying rotating machine faults as described in [7] and the detection of tool wear[8]. Adaptive signal processing for detecting bearing signal corrupted by severe background noise, such as adaptive noise cancellation (ANC), has been shown to have potential but requires two inputs.

Recently other signal processing techniques to enhance an internal defect have been introduced, such as auto-regressive techniques [9,10] and application of higher order moment spectral analysis [11,12]. Although auto-regressive techniques have the potential for early detection of component defects, the complexity of these techniques impedes their implementation. The authors in [10] concluded that they should not replace conventional methods but supplement them.

Blind deconvolution (equalization) is a technique used to recover the desired signal from a single received channel without any *a priory* knowledge about the unknown channel. The technique has been widely used in network communication since the 1980s [13]. In the past 20 years numerous techniques have been proposed to enhance signals based on single inputs. However, an improved Blind Deconvolution technique has been used in in a variety of applications [14-16]. A major advantage of blind deconvolution is that it does not require a training stage, which is essential in conventional equalization [17,18]. Recently the authors have successfully applied the technique to recover fault information from corrupted bearing signals [18a, 18b].

In this paper higher order solutions to blind deconvolution using the EVA algorithm developed in [19] was applied to enhance bearing signals corrupted by a variety of periodic noise. The work presented in [20,21] was use to analyse the gain response of the equalizer. Both the Modified Crest Factor (MCF) and the Arithmetic Mean (AM) in conjunction with a neural network was used to optimise filter length of the equaliser. The results show that the technique was successful in enhancing the defective bearing signal corrupted by periodic noise and the MCF and AM can be used to optimise the filter length.

2 BLIND DECONVOLUTION THEORY

In practical situations, the observed signal $y[n]$ from a measurement is not the original input signal, but is the convolution of the original input signal $x[n]$ with the impulse response system $h[n]$ plus a corrupting noise $s[n]$ as shown schematically in Fig 1. The original input signal is inaccessible and the impulse response system is unknown.

Deconvolution refers to the determination of the impulse response of the system $h[n]$ or determination of the input signal $x[n]$ where the output of the system $y[n]$ is typically accessible and the knowledge of the input signal $x[n]$ is unavailable.

The observed signal from measurement is given by:

$$y[n] = h[n] * x[n] + s| \qquad (1)$$

$$y[n] = \sum_{m=-\infty}^{+\infty} h[m]x[n-m] + s[n] \qquad (2)$$

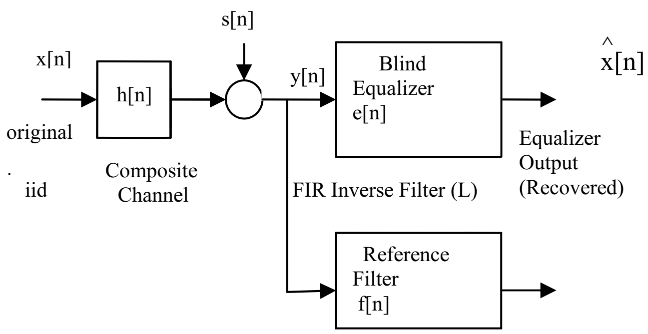


Fig 1. Schematic of observed signal and the Eigenvector Approach (EVA)

In order to reconstruct the original signal, a system that optimally removes the corruption from the unknown channels is needed. This system is known as an equalizer in signal communication. Unlike the ordinary equalizer, the blind equalizer performs without access to the original input or any known training sequence, target or desired signal as in adaptive filtering.

The output of the blind equalizer is denoted by $\hat{x}[n]$:

$$\hat{x}[n] = e[n] * y[n] \qquad (3)$$

$$\hat{x}[n] = \sum_{m=0}^{L} e[m]y[n-m] \qquad (4)$$

The objective is to find the Finite Impulse Response (FIR) of the inverse filter $e[n]$ that will give $\hat{x}[n]$ is as close as possible to the delayed original signal $x[n-k_0]$ in the Mean Square Error (MSE) sense [19] :

$$MSE(e,k_0) \underline{\underline{\Delta}} \sum \{| \hat{x}[n] - x[n-k_0] |\}^2 \tag{5}$$

The eigenvector algorithm (EVA) method used by the Kammyar and Jellonak [19] which is based on the cross kurtosis of two higher order cumulants and maximizes the kurtosis criterion which is applied to determine the impulse response of the equalizer. Using this method, the characteristic of the inverse channel can be estimated directly from a number of samples of the observed signal. Figure 1 shows the schematic of the Eigenvector Approach (EVA). An equalizer, e[n], is introduced to the channel output y[n] to produce an output signal $\hat{x}[n]$. The second inverse filter with the same order is introduced as a reference system filter. The task of the second filter is to generate an implicit sequence of reference data for the iteration process. The impulse response of second filter is denoted by f[n].

The solution to blind deconvolution is based on the maximum *"cross kurtosis"* quality function. Cross kurtosis between $\hat{x}[n]$ and u[n] can be used as a measure of equalization quality. Using the convolution between the equalizer input and output signals, the cross-cumulant of the observed signal y[n] and the reference signal u[n], can be used to maximize the quality function.

$$\text{Maximize } \left| e^* C_4^{uy} e \right| \text{ subject to } e^* R_{yy} e = \sigma_d^2 \tag{6}$$

where C_4^{uy} represents the Hermitian cross cumulant matrix of the output signal u[n] and observed signal y[n]; R_{yy} represents the Toeplitz autocorrelation matrix of the observed signal y[n] and e is the equalizer impulse response. σ_d^2 denotes variance of the signal and " * " denotes the convolution operator.

The qualifying function equation (6) is a quadratic in the equalizer coefficient and can be expressed as follows

$$C_4^{uy} \; e_{EVA} = \lambda \; R_{yy} \; e_{EVA} \tag{7}$$

Equation (7) is known as EVA equation. The coefficient vector e_{EVA} is obtained by choosing the eigenvector of $R_{yy}^{-1} \; C_4^{uy}$ associated with the maximum magnitude eigenvector λ.

3 DETERMINATION OF OPTIMUM FILTER LENGTH

The filter length of the blind equalizer plays a vital role in the deconvolution process due to different operating conditions and types of bearing faults. It is generally know that when a bearing has fatigue crack, the vibration signal will generate a series of periodic impulses. The measurement of energy of the recovered signal can be used to determine the condition of bearing and a high value of RMS above a certain threshold normally indicates the extent of the damage. However, in some cases this criterion has had limited success.

The Crest Factor is defined as a ratio of the maximum peak of the signal over the RMS value and can be used as a feature to identify a bearing fault. In this study a modified Crest Factor, which is the average of peaks over the RMS value, was used to optimise the equaliser length. The modified Crest Factor can be expressed as:

$$\text{Modified Crest Factor} = \frac{\text{Average of Maximum Peaks}}{\text{RMS}} = \frac{\frac{1}{M}\sum_{i=1}^{M} X_{p_j}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} X_i^2}} \tag{8}$$

where X_{P_j} corresponds to the amplitude of vibration signal exceeding a certain threshold; M is number of sequences exceeding the threshold; x_i stands for the signal amplitude and N is the total number of sequences in a signal.

The overall vibration level of a faulty bearing increases when the fault advances. A simple parameter which can quantify the difference between the spectra is the Arithmetic Mean (AM) and is defined as:

$$AM = 20\mathrm{Log}\left(\frac{\frac{1}{N}\sum_{i=1}^{N} A_i}{10^{-5}}\right) \qquad (9)$$

where A_i are Fourier coefficients of the vibration spectrum. The change in the value of AM with succeeding spectra gives a measure of the change in spectra.

In this paper these two criteria were used to determine the filter length of the equalizer. A neural network (NN) was incorporated to train the behaviour of the system which generates the optimum filter length. Several damaged bearings with outer race defects, inner race defects and ball defects with varying defect sizes were used as target sets for the NN to automatically determine the optimum filter length for the equalizer. A data training set of the neural network was obtained with different defect sizes and by varying the speed of the shaft from 600 RPM to 1800 RPM. Four vibration parameters were considered as the input parameters to the neural network namely; crest factor, kurtosis, arithmetic mean and the Matched filter RMS (Mfrms) of the observed signal [11]. The target of the neural network was to determine the optimum filter length.

4 RESULTS AND DISCUSSION

4.1 Simulation Studies

The robustness of *blind deconvolution* in enhancing the bearing signal was initially demonstrated by corrupting the observed signal with a range simulated deterministic noises. In this test a pre-recorded damaged bearing signal (fig 2(a)) was corrupted by periodic sinusoidal noise to form the observed signal. To simulate a real situation, noise was generated in the laboratory using a deterministic noise generator. The amplitude and the frequency of the generated noise can be varied to produce different signal-to-noise-ratios. The lowest SNR created was -49.68 dB resulting in the bearing signal being totally masked by the noise as shown in Fig 2(b). For higher SNRs, the signal impacts were more prominent and were not considered in this research.

Figure 2(c) shows the output after the BD process and it clearly demonstrates that this technique has the capability of enhancing the bearing signal even with a very low SNR (-49.68 dB). In this test, the filter length was determined using a trial-and-error approach.

To observe the removal of all corrupting noise components, the gain response of the equalizer was plotted as shown in Fig 3. It can be seen that this equalizer acts like a notch filter in removing the corrupting periodic frequencies at 250, 500, 1000, 2000 and 4000 Hz. It can be seen that the BD process removes all periodic components of those frequencies plus all other inherent periodic components.

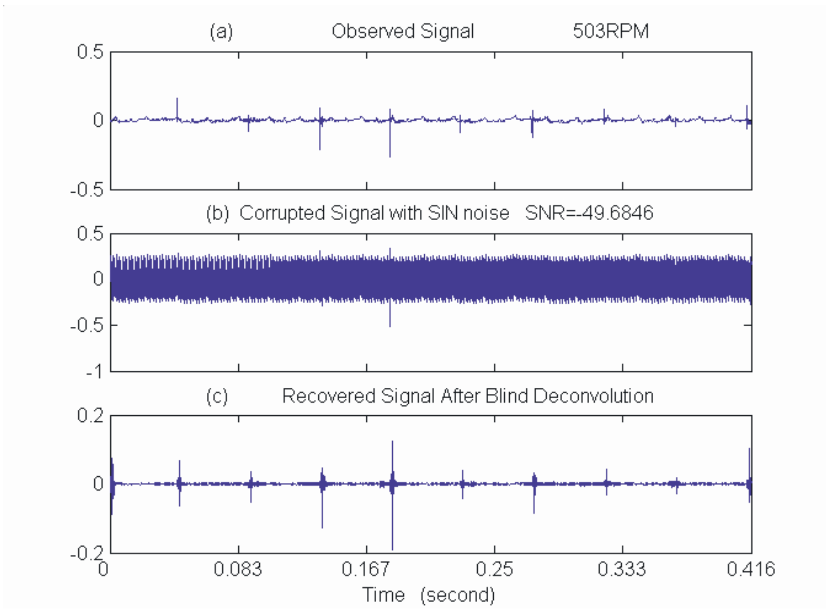


Fig. 2 (a) Original damaged bearing signal; (b) Observed corrupted signal with 5 periodic noises, 250, 500, 1000, 2000 and 4000Hz: and (c) Recovered bearing signal

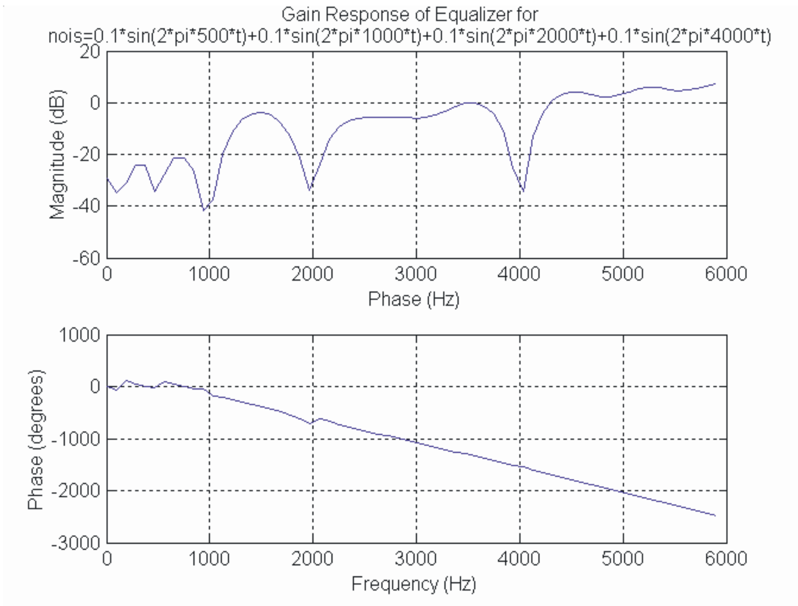


Fig 3. Gain Response and phase plots of an equalizer.

4.2 Optimum filter length

The results above show that BD can remove periodic noise and that the equaliser behaves like a notch filter. In real life applications, the method of trial-and-error approach to determine the optimum filter length for the equalizer is not viable. The above results demonstrated that with different S/N ratios, damaged sizes and operating speeds, different filter lengths were required for a successful application.

The results of optimum filter length trained using a Neural Network with defect sizes of 0.1 to 0.5 mm, and located in different bearing elements with varying speed from 600 RPM to 1800 RPM are shown in Table 1. Only those with a simulated defect on the outer race are presented for discussion.

Table 1. Neural Network trained optimum filter length for the equalizer

	Damage size (mm)	600RPM	1200RPM	1800RPM
Outer race	*0.1*	7.00E+01	4.60E+01	5.60E+01
	0.2	3.60E+01	6.840E+01	5.40E+01
	0.5	6.80E+01	6.800E+01	5.20E+01

From the table it is not possible to establish a consistent trend of equalizer filter length against the operating speeds or defect sizes. However, for high operating speeds, the filter lengths for an outer race defect are between 52 and 56. At lower speeds (600 rpm) the filter lengths fluctuate between 36 and 70. Similar trends were obtained for damage on the inner race and ball.

The optimum filter length was selected based on the modified crest factor and arithmetic mean criterion. Figure 4 shows a typical plot of CF and AM versus filter length for an outer race fault of 0.1 mm defect width with rotational speed of 600 RPM. The algorithm selects the optimum filter length based on the values of CF and AM. It can be observed from the plot that the CF values stabilized at around 70 filter length and remained fairly constant all the way to 240 filter length. In the AM plot it can be seen that the AM amplitudes decrease from -35 dB to -50 dB until 70 filter length. Beyond that value, the AM amplitude remained fairly constant at about -50 dB. The optimum filter length of 70 for this particular case was then input to a data training set for the equalization process.

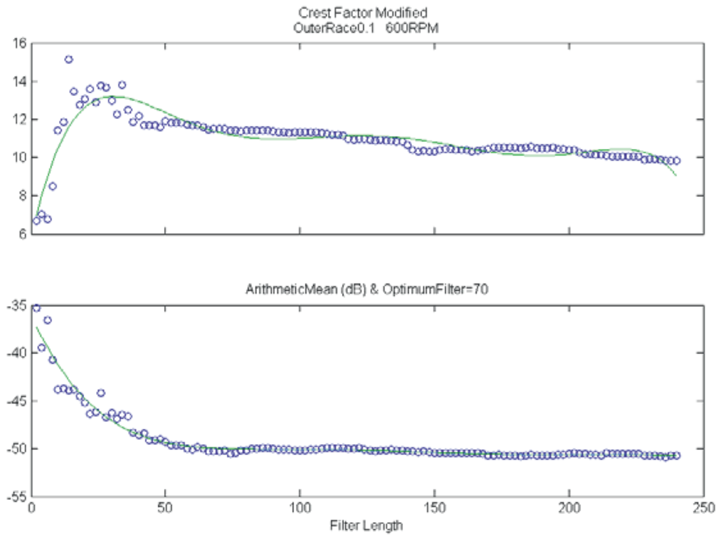


Fig 4. CF and AM plot for Outer Race Defect of 0.1 mm width at 600 RPM

4.3 Removal of corrupting noise

Using the results of optimum filter length, an observed signal with an outer race defect of 0.1 mm operating at 600 rpm was put to the test to see if the NN determined filter length can be used to optimise the blind equalizer. Figures 5(a) and 5(b) show the original observed bearing signal of an outer race defect of 0.1mm and the corrupted bearing signal by a sinusoidal noise of 500Hz, respectively.

It can be seen from Fig 5(b) that the observed signal is totally masked by the corrupting noise. With a filter length of 70, the recovered signal from the BD process is shown in Fig 5(c). It can be seen that blind deconvolution with optimum filter length has enhanced the original bearing signal and has eliminated the background noise. The impulses from the defect on the outer race can be clearly seen and the SNR has been improved.

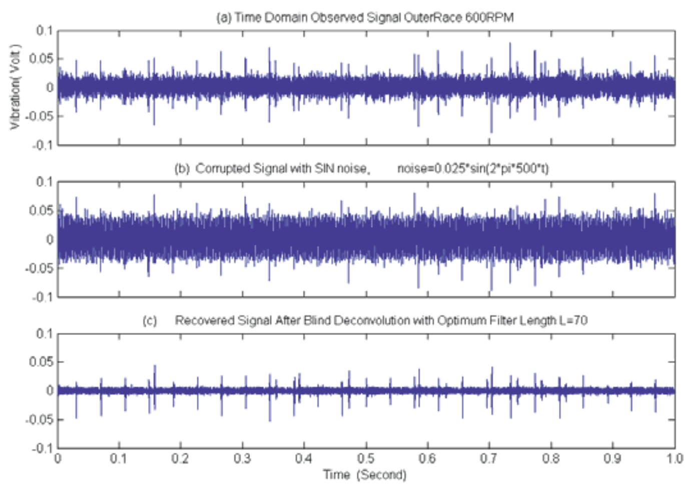


Fig 5. (a) Observed signal at 600 rpm with an outer race defect of 0.1mm; (b) Observed signal corrupted by periodic noise and (c) Recovered signal after the BD algorithm.

The recovered signal shows a consistent impulsive signal of a damaged bearing. The corrupted signal has a kurtosis value of 2.78 which indicates a healthy bearing and can be misleading as the damaged signal is corrupted by high level noise. The recovered signal shows the impulsive nature of a damaged bearing with a kurtosis value of 9.06. The results show that it is impossible to notice a damaged bearing signal from the corrupted signals as the Kurtosis value indicates a good bearing.

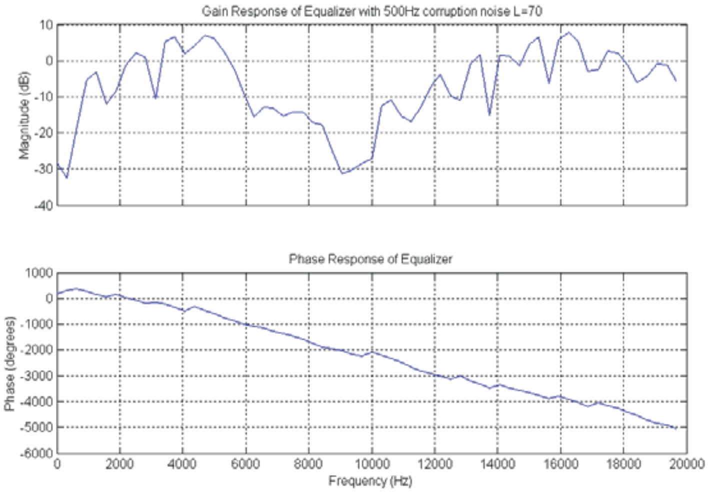


Fig 6. Gain response and phase response of the equalizer of an outer race defect of 0.1mm.

In order to understand the nature of the equalizer in the blind deconvolution algorithm, the gain response of the equalizer was plotted using the computed coefficient of the equalizer e[n]. Figure 6 shows the gain response of the equalizer (filter) which was used in this experiment. It can be seen from the gain response plot that there is a notch at the corrupting frequency of 500 Hz with -34dB and a band of frequencies from 6200 to 11500 Hz. Although there was no prior knowledge of the simulated noise it can be seen that the equalizer has successfully eliminated the sinusoidal noise and a band of high frequency signals. The resonant frequencies of the defective roller bearing were found by performing an impact test on the ball bearing. It was found that one of the excited frequencies of the bearing's structure was around 8000 Hz.

5 CONCLUSIONS

This study has revealed the advantages of blind deconvolution as a technique to recover the original signal of a faulty bearing corrupted by noise and distorted by the transmission path. Since characteristic frequencies contain very little energy, and usually are overwhelmed by noise and higher levels of macro-structure vibration, it is difficult to identify them in the time domain. The procedure to obtain the optimum inverse filter is addressed, considering the influences of input parameters of blind deconvolution. The modified Crest Factor (CF) and Arithmetic Mean (AM) graphs was used successfully to to determine the optimum filter length. It was found that there is a good correlation between both graphs to determine the optimum filter length. Using a slope and finite difference programs, the optimum filter length of the equalizer can be identified and applied to the blind deconvolution algorithm automatically. This study has revealed that blind deconvolution behaves like a notch filter in eliminating periodic noise. The results show that the proposed algorithm works very well with neural networks to determine the optimum filter length.

6 REFERENCES

[1] Morando, L.E., “Bearing condition monitoring by the shock pulse method”, Iron and Steel Engineers, pp 40-43, Dec 1998

[2] Li, J.C., Ma, J. and Hwang, B., “Bearing localized defect detection by bicoherence analysis of vibrations’, Tran of ASME, J Eng Ind, vol 117, 625-629, 1995.

[3] Tandon, N, “A comparison of some vibration parameters for the condition monitoring of rolling element bearings”, Measurement, Vol 12, pp 285-289, 1994.

[4] Su, Y.T., and Lin, S.J., “On initial fault detection of a tapered rolling bearing: frequency domain analysis”, J Sound and Vib, Vol 155, 78-84,1992.

[5] Chaturvedi, G. and Thomas D.W, “Bearing fault detection using adaptive noise cancelling.”, Trans of ASME 280, vol 104, 1982.

[6] Tan, C.C., and Dawson, B., “An adaptive noise cancellation technique for the enhancement of vibration signatures from rolling element bearings”, On-Line monitoring of continuous process plants, Pub. Soc of Chemical Industries, Ellis Horwood, 1981.

[7] Lee, S.K., and White, P.R., “Fault identification for rotating machinery using adaptive signal processing and time-frequency analysis”, ASME Des Eng Tech conf, 16^{th} Biennial Conf on Vib and Noise, DETC97/VIB-4236, 1997.

[8] Tan, C.C. & Okada, Y., ”Detection of rolling bearing signal corrupted by noise of similar frequency using ANC”, Proc of Info, Comm & S.P , ICICS’97, Singapore, Aug 1997

[9] Bailie, D.C. and Mathew, J., “A comparison of autoregressive modelling techniques for fault diagnosis of rolling element bearings”, Mechanical Systems and Signal Processing, Vol 2, 1-17, 1996.

[10] Dron, J.P., Rasolofondraibe, L., Couet, C. and Pavan, A., “Fault detection and monitoring of a ball bearing benchtest and a production machine via autoreressive spectrum analysis”, J of Sound and Vib, Vol 218(3), 501-525, 1998.

[11] Zoubir, A.M. and Bohme, J.F.,”Application of higher-order spectra to the analysis and detection of knock in combustion engines”, Higher-order Statistical Signal Processing (Editors B. Boashash et.al), London Longman, 1995.

[12] Lee, S.K. and White, P.R.,”Fault diagnosis of rotating machine using Wigner higher order moment spectra”, Mechanical Systems and Signal Processing, vol 11(4), 637-650, 1997.

[13] Godard, D.N., “Self-recovering equilization and carrier tracking in two-dimensional data communication systems”, IEEE Trans. Comm., **28**, 1867-1875, 1980.

[14] Nandi, A.K., Mampel, D. and Roscher, B., “Blind deconvolution of untrasonic signals in non destructive testing”, IEEE Trans., on Sig. Proc., **45**(5),1382-1393, 1997.

[14] Zheng, F.C. and McLaughlin, S., “Blind equilization on nonminimum phase channels: Higher order cumulant based algorithm”, IEEE Trans. Sig. Proc.,**41**(2), 1993.

[16] Feng, C.C. and Chi, C.Y., “Performance of cummulant based inverse filters for blind deconvolution”, IEEE Trans. On Sig Proc., **47**(7), 1999.

[17] Haykin, Simon, “Adaptive Filter Theory”, 3^{rd} edition, Chapter 18, Prentice-Hall International, INC., pp 772-814, 1992

[18] Haykin, Simon, “Blind Deconvolution”, Prentice-Hall International, INC., 1994

[18a] Tan ACC, Nibito W and Mathew J, “Detection of incipient rolling element bearing failure by blind deconvolution”, Proc of the APVC 2001, pp 25-29, Hangzhou, China 2001

[18b] Tan ACC and Mathew J, “The adaptive noise cancellation and blind deconvolution technique for detection of Rolling element faults – A comparison”, Proc of the 3rd ACSIM, Cairns, Sept 2002

[19] Boss, D., Jelonnek, B., Kammeyer, K. D., “Eigenvector Algorithm for Blind MA System Identification”, ELSEVIER Signal Processing, Vol. 66, No. 1, April 1998.

[20] Karimi, M, Mathew, J., Tan, ACC., Senadji, B., “Bearing fault detection using blind deconvolution and time interval averaging”, Proceeding of the 3rd Asia Pacific Conference, 2003

[21] Karimi, M, Tan, ACC., Mathew, J., Senadji, B., “ Blind deconvolution as a notch filter in suppressing correlated noise”, Proc of VETOMAC – 3 and ACSIM 2004, 484-491, New Delhi, 2004

BLIND EQUALIZATION BASED EIGENVECTOR ALGORITHM FOR THE RECOVERY OF MECHANICAL VIBRATIONS

Peter W. Tse, Szymon W. Gontarz and Xiaojuan Wang

Smart Engineering Asset Management Laboratory (SEAM),
Department of Manufacturing Engineering & Engineering Management,
City University of Hong Kong, Hong Kong, China

Abstract: Many advanced techniques have been developed for the analysis of mechanical vibrations. It is one of the prerequisites to vibration-based machine fault diagnosis that the vibration signal measured from a machine component must be well isolated from other vibration signals that are generated by adjacent components. Due to the physical constraints of installing sensors in the machine, sometimes only one sensor can be installed. Hence, the sensor will collect an aggregated source of vibrations rather than just the vibration generated from the inspected component. Manufacturing machines are prone to such interference of multiple vibrations. Thus the fault-related vibration must be recovered from the aggregated sources for accurate fault diagnosis. In this paper, the eigenvector algorithm (EVA) of blind equalization (BE) is applied to the recovery of mechanical vibration signals. The conventional EVA can extract only one dominant source from the collected data at a time. In this paper, we propose an enhance EVA that is constructed with the method of channel extension and further post-processing algorithm to recover multiple sources of vibrations. That is, besides the dominant vibration, other less dominant vibrations but relevant to existing faults can also be recovered by using the analyses of correlation and kurtosis. The experiments performed on real vibrations that are generated by industrial machines present the effectiveness of the proposed methods.

Key Words: Blind equalization; Blind deconvolution; Vibration analysis; Machine fault diagnosis

1 INTRODUCTION

Rotational components in machine are often related to the use of bearings and shafts, any faults of which will result in the ineffiecient operation or breakdown of the machine. The additional vibration that is generated by the defects can provide useful information for the fault diagnosis. Currently there are many methods for this aim, but they are ineffective in practical applications if the vibration of interest interferes with other source signals seriously. In the machine with numerous small components closely packed, an inspected source is easily contaminated by vibrations or noises from other adjacent components, which means the signal that is acquired by a sensor is usually the superposition of a number of sources. Hence, the recovery of concerned signals of interest is the prerequisite to apply the vibration-based analysis methods for sucessful fault diagnosis.

A great deal of efforts have been paid for the approaches of vibration recovery in mechanical field recently, which mainly include exact wavelet analysis [1], principal component analysis [2], neural networks [3] and adaptive noise cancellation [4]. However, their effectiveness in extracting fault-related vibration is not always satisfactory, especially when the fault symptom is small whereas the interference or noise is relatively large. Therefore, more efficient methods of vibration recovery are required for an accurate and reliable diagnosis of rotational components. Recently, blind equalization (BE), which was initially used to recover sound signal, has been successfully applied to engineering problems in many other industial fields. The fundamental idea of BE is to derive from the received signal the equalizer characteristic without any prior knowledge of the sources and propagation channels. It can recover the source with one sensor even when time delay is involved in the signal collection. Another advantage of BE is that it does not require a training process, which is necessary in many conventional signal equalization methods. EVA is one of the most popular solutions of BE, and it turns out to be very useful in recovering the fault-related impacts. Jelonnek et al. [5] applied EVA to get the optimum result in the received signal that is an aggregation of data from many unknown transmission channels. However, as a serious limitation for its appliation in machine fault diagnosis, EVA can recover only one dominant source from the mixture that is collected by a single sensor.

In this paper, the eigenvector algorithm (EVA) of blind equalization (BE) is applied to the recovery of mechanical vibration signals. We propose an enhanced EVA that consists of channel extension and a post-processing method to recover multiple sources of vibrations. The post-processing method is based on correlation and kurtosis as the criteria to determine the relation between recovered signals and inpected components. The real vibrations that are generated by industrial machines are used to verify the effectiveness of the proposed methods. The results show the characteristic difference between conventional and enhanced EVA in recovering fault-related vibrations from mixed sources, which indicates that the enhanced EVA can be taken as a promising method for the recovery of vibrations in machine fault diagnosis.

2 CONVENTIONAL EVA SOLUTION TO BLIND EQUALIZATION

One of the solutions for linear blind equalization was found and expressed as the theory related with eigenvector in 1994 [6], namely EVA, and it has been the most popular technique of BE nowadays. The complete model of EVA is introduced in Figure 1. EVA takes over all assumptions from blind equalization theory. Moreover, it applies an additional virtual equalizer *f(k)*, which has the same order as equalizer *e(k)* in conventional BE, to generate an implicit sequence as reference data for estimation of reference output *z(k)* in subsequent iterative process.

$$z(k) = x(k) * f(k) \tag{1}$$

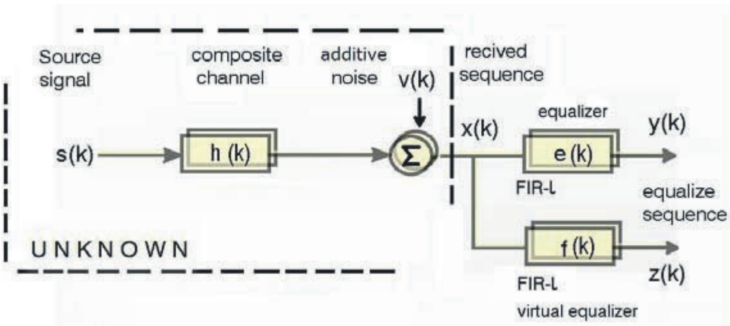


Figure 1. The conventional configuration of EVA.

The EVA requires the two-dimensional fourth order cross-cumulant matrix for equalizer between the samples of the received signal *x(k)* and the reference system output *z(k)*. The cross-cumulant matrix with size $(l+1)\times(l+1)$ is derived as

$$C_4^{zx} = \mathrm{E}\left\{|z(k)|^2 xx^*\right\} - \mathrm{E}\left\{|z(k)|^2\right\}\mathrm{E}\left\{xx^*\right\} - \mathrm{E}\left\{xz(k)\right\}\mathrm{E}\left\{x^*z^*(k)\right\} - \mathrm{E}\left\{xz^*(k)\right\}\mathrm{E}\left\{x^*z(k)\right\} \tag{2}$$

where x and x^* denote the vectors $[x^*(k), x^*(k\text{-}1), \ldots, x^*(k\text{-}l)]^T$ and $[x(k), x(k\text{-}1), \ldots, x(k\text{-}l)]$.

Using the autocorrelation matrix of received data *x(k)* and the vector of the equalizer impulse response, we finally get $\left|e^* C_4^{zx} e\right|$ as the criterion. The proper adjustment of equalizer coefficients could be achieved by maximizing it on the condition that

$$E(x^*x) = E\{[(v^*e)^*](v^*e)\} = E(e^*vv^*e) = e^*E(vv^*)e = e^*Re, \quad \text{so } |e^*R_{xx}e| = \sigma^2 = const. \tag{3}$$

The EVA technique was discussed as a promising tool for mechanical fault diagnosis in many literatures (see [7, 8]). Some researches have proved that the EVA as a solution of BE is effective in recovering the periodic or impulsive type of signal, which appears mostly in faulty machines, from interferences and noises with only one sensor. However, there are still many shortcomings that are associated with the application of EVA in mechanical applications. The signals in manufacturing machines are always very complex and prone to the interference from various adjacent components and noises. It means that the signal that is collected with one sensor is definitely a mixture of numerous sources. Although the conventional EVA can recover the dominant source of the aggregated signal that has the highest amplitude, other fault-related sources with the lower amplitudes may not be able to be recovered.

3 THE ENHANCEMENT OF EVA FOR THE RECOVERY OF MECHANICAL VIBRATIONS

Enhancement of the conventional EVA is proposed to facilitate the recovery of multiple mechanical vibrations, which may not only include the primary source but also some other subdominant sources in the collected signal. The practice of proposed method in this stage of research mainly involves bearing diagnosis. Our enhanced EVA is constructed by the methods of channel extension and post-processing algorithm. Fractional Tap Spacing (FTS) equalizer is also considered to improve the performance of proposed EVA method. The advantages of FTS have been discussed in some literatures (see [9]), and detailed description about FTS equalizer can refer to the work of Jelonnek et al. [5].

According to the concept of FTS equalizer, the multirate baseband model of BE with spaced equalizer can be decomposed and extended to multiple channels that are showed in Figure 2. That is, the EVA can be applied to $x_1(k)\ldots x_M(k)$, where M denotes the factor of up-sampling or down-sampling in FTS [5]. Compared to the model of conventional EVA, such an enhanced EVA possesses more than one parallel channels, e.g. $h^o(n)$ … $h^e(n)$, which refer to different propagation paths and correspond to the source signals that are generated in the inspected machine. In order to recover original signals, the equalizer and reference system must be consequently divided into the same number of channels [10]: $e^o(n)$... $e^e(n)$ and $f^o(n)$... $f^e(n)$, each of which is characterized by specific order of filter l_j. Total number of equalizer coefficients will be finally adjusted by EVA amounts

$$\tilde{l}+1=\sum_{i=1}^{M}(l_i+1)=M+\sum_{i} l_{i.} \tag{4}$$

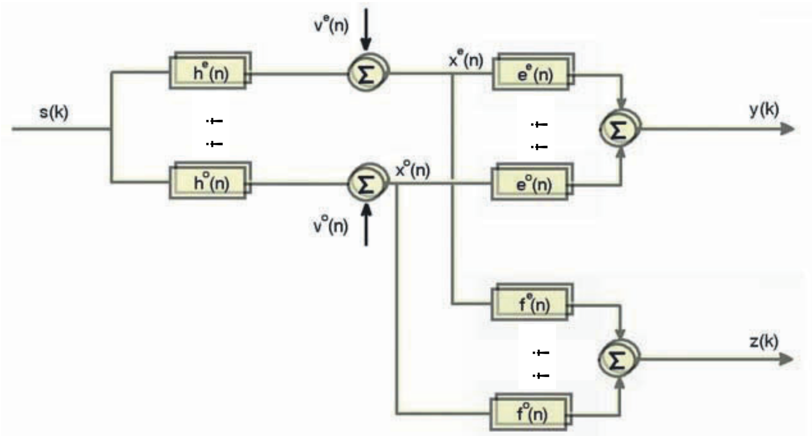


Figure 2. The enhanced EVA for multirate baseband model with multiple transmission channels.

In order to obtain an effective solution for the enhanced EVA, $q(k)$ and $r(k)$ are defined as the overall system where the input is $s(k)$ and the outputs are $y(k)$ and $z(k)$. The criterion based on fourth order cumulants between $y(k)$ and $z(k)$ is used to identify the quality of equalization [11]. Output $y(k)$ can be expressed as a sum of input sequences $x(k)$ of the equalizers

$$y(k)=\sum_{i=1}^{M} x_i(n) * e_i(n)=\tilde{x}_n^{*}\,\tilde{e}\,, \tag{5}$$

where the vectors $\tilde{x}_n$ and $\tilde{e}$ with a length of l+1 are composed of M subvectors.

$$\tilde{x}_n^{*} = \left[x^{*}_{n,1}, x^{*}_{n,2}, \ldots, x^{*}_{n,M}\right]^T \quad \text{where} \quad x^{*}_{n,i}=\left[x_i(n)\ldots,x_i(n-l_i)\right] \tag{6}$$

$$\tilde{e}=\left[e_1^T, e_2^T, \ldots, e_M^T\right]^T \quad \text{where} \quad e_i=\left[e_i(0),\ldots,e_i(l_i)\right]^T \tag{7}$$

Then quality criterion of cross kurtosis is applied to $y(k)$ that is obtained in equation (5),

$$\tilde{C}_4^{zx} = \mathrm{E}\left\{|z(k)|^2\,\tilde{x}_n\,\tilde{x}_n^{*}\right\} - \mathrm{E}\left\{|z(k)|^2\right\}\mathrm{E}\left\{\tilde{x}_n\,\tilde{x}_n^{*}\right\} - \mathrm{E}\left\{\tilde{x}_n\,z(k)\right\}\mathrm{E}\left\{\tilde{x}_n^{*}\,z^{*}(k)\right\} - \mathrm{E}\left\{\tilde{x}_n\,z^{*}(k)\right\}\mathrm{E}\left\{\tilde{x}_n^{*}\,z(k)\right\}, \tag{8}$$

$$\text{maximize } |\tilde{e}^{*}\,\tilde{C}_4^{zx}\,\tilde{e}|, \qquad \text{subject to } |\tilde{e}^{*}\,\tilde{R}_{xx}\,\tilde{e}|=\sigma^2=const., \text{ where } \tilde{R}_{xx}=E\{\tilde{x}_n\,\tilde{x}_n^{*}\} \tag{9}$$

The optimization of this equation leads to the generalized eigenvector problem ($C_4^{zx}e_{EVA}=\lambda R_{xx}e_{EVA}$). Its solution is obtained by choosing the eigenvector with the largest eigenvalue. In this way coefficient vector $\tilde{e}_{EVA}=\left[e_{EVA,i}^T,\ldots,e_{EVA,M}^T\right]^T$ is acquired and applied to blind equalization in FTS configuration with two channels. There is different sequence of filter coefficients $e_{EVA,i}=\left[e_{EVA,i}(0),\ldots,e_{EVA,M}(l_i)\right]^T$ (i=1,…, M) for each channel. In the enhanced EVA solution, the impulse response of equalizer $\tilde{e}(n)=e(k)$ is constructed by interleaving the $e_{EVA,\,i}$ components.

To obtain multiple outputs in the equalization process, different sets of filter coefficients are proposed. The filters in proposed method are set with different lengths. In the example of this paper, the enhanced EVA will adjust the number of coefficients to 4, 8, 16, 32, 64 for $L=2^i$ (i=2, 3…6), so that five different characteristic of blind equalizers are obtained. After applying these filters to inspected signals

$$x(k) * e_i(k) = y_i(k), \tag{10}$$

the results of the equalized signals $y_i(k)$ are obtained, each of which may relate to a different vibration component in the aggregated raw input signal.

After obtaining the multiple equalized signals based on the conventional EVA, first, our post-processing method will calculate the correlation coefficients (*cc*) between the collected raw vibration signal x(t) and each of the obtained equalized signals yi(k). The signal with the largest value of *cc* in existing equalized results is found for an identification process, and then it will be converted to its frequency domain. From its frequency spectrum, we will verify the existence of bearing characteristic frequencies (BCFs) to confirm whether this signal is generated from a bearing. If it is belonged to the desired bearing vibration and the sensor is mounted on the bearing's housing, then the process is halted. If not, then it is necessary to search the bearing vibration from the rest of the equalized signals. Second, to avoid sequential searching just based on *cc* that is time intensive, a fast searching approach is introduced here by using Kurtosis. Kurtosis is a statistical indicator to evaluate the distribution of impulsive characteristic of a vibration signal, and it is particularly suited to monitor bearing's faults as most of them will exhibit impulsive characteristic in its vibration signal if the bearing is defective. The value of kurtosis for each remained equalized signals will be calculated. The equalized signal that has the largest kurtosis will then be identified. If the signal belongs to one of the BCFs, then the process is halted, whereas a further step will be processed. The identification process will continue if the last step has failed to identify the bearing vibration. The remained equalized signals will be ranked again according to their *cc* or kurtosis and taken the verification of BCFs. If it is the bearing vibration, the searching process is halted; if not, then the process will be repeated until the bearing vibration can be identified or all of the equalized signals are investigated. Note that if an equalized signal that may have a large kurtosis value but a low *cc* value, and cannot be verified as one of the BCFs, is probably noise. With the help of the enhanced EVA, both the dominant vibration and the interested bearing vibration that has been overwhelmed by other larger vibrations can be recovered. Some useful signals relevant to the components of the inspected machine can also be extracted in this process for further identification. A real case of recovering such overwhelmed vibration using the enhanced EVA is presented in the following section.

4 RECOVERY RESULTS OF REAL SIGNAL IN INDUSTRIAL MACHINE

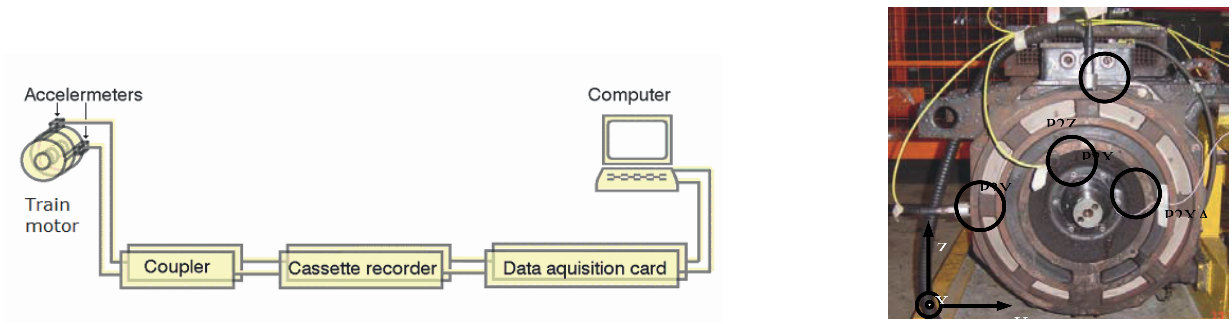


Figure 3. The schema of the testing system, and the position of four accelerometers in the traction motor.

In our previous work [4], two fault-related vibrations that were generated by two faulty components of a traction motor were obtained from the data collections of multiple sensors using the method of adaptive noise cancellation (ANC). The diagnosis of the two faulty components was confirmed by the maintenance staff after an overhaul had been done on the traction motor. To investigate the effectiveness of the conventional EVA and the enhanced EVA in the recovery of vibration signals, the experiments were conducted on the same traction motor using the same collected signals for comparison purpose.

The real vibration signals from a traction motor in a subway train were simultaneously collected for fault diagnosis to validate the proposed method. The engine was uncoupled from the gearbox for the experiments. The motor consisted of a 250 kg rotor that was supported by two rolling element bearings, one of which was located on the drive end and the other on the non-drive end of the motor. The vibrations in three orthogonal directions of each bearing were respectively monitored by three accelerometers at the positions P2X, P2Y and P2Z. An additional sensor labeled as P2XA was also mounted at the axial direction, but much closer to the bearing house of the motor than the other sensors. The experiment schema with all components that were involved in the testing is shown in Figure 3. The signals from the sensors were amplified by a coupler (Kistler 5134), and a digital cassette recorder (Sony PC204Ax) was used to acquire the signal simultaneously at the sampling rate of 48k Hz for each of channels. The record length of data was approximately 1.5 minutes. Then the signals were transmitted to PC with a data acquisition card for further processing. The sampling rate of data acquisition was 32.8k Hz.

The conventional EVA and the enhanced EVA in this research were respectively applied to the testing of the motor. Both accelerometers P2XA and P2X could acquire the vibrations from the motor and the bearing. However, the sensor P2XA was mounted closer to the bearing house, and so it would capture more vibration that was generated by the bearing than the sensor P2X. Therefore, the vibration that was obtained at the position of P2XA was analyzed to recover the overwhelmed bearing vibration from all of motor signals. Figure 4 shows the raw vibration from P2XA and the result after the application of the conventional EVA.

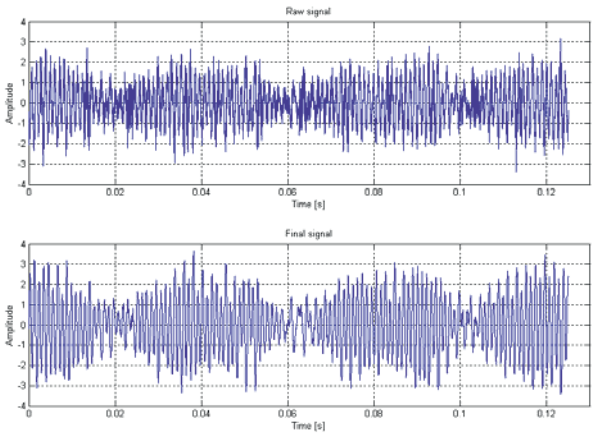


Figure 4. Raw signal and recovered signal after implementation of the conventional EVA.

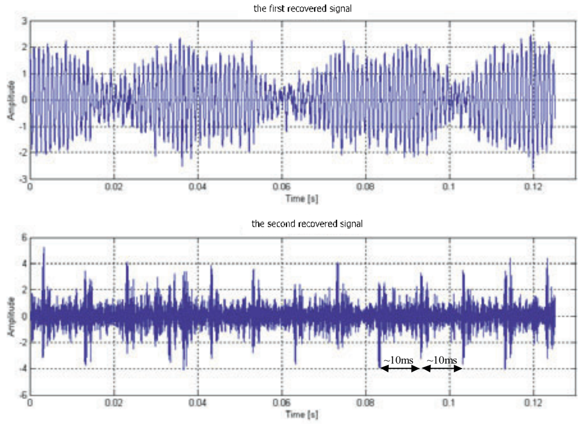


Figure 5. Two recovered signals after implementation of the enhanced EVA.

There were two faults existing in the tested motor, one of which was a motor eccentric problem due to poor workmanship on the clearance between the rotor and the stator. The second fault was an outer-race defect occurred in the inspected bearing. The raw signal collected in the top diagram contains the fault-related vibrations that were generated by both the bearing and the motor. The bearing signal is difficult to be identified because it was overwhelmed by the dominant vibrations from the faulty motor. The conventional EVA adjusted the equalizer length to 8 to recover the motor vibration because of its dominant status, which is shown in the bottom diagram of Figure 4. However, the bearing vibration still remained ambiguous. The proposed enhanced EVA shows more efficient ability than the conventional EVA in the recovery of mechanical vibrations. The results of this method are presented in Figure 5. As proposed in section 3, the motor vibration was firstly recovered as a dominant component of the collected mixture as shown in the top diagram of Figure 5. A filter with the length of 8 was used in the experiment to obtain this result. The recovered vibration signal had a correlation coefficient of approximately 0.8, which was the largest value in all equalized results. Besides the motor vibration that can also be recovered by the method of conventional EVA, the enhanced EVA enables the recovery of the second signal component, which is shown in the bottom diagram of Figure 5. The signal recovered by the equalizer with a length of 32 has a largest kurtosis in all equalized results, and was marked as the second dominant signal. In addition, its impulsive characteristics further proved that the second recovered vibration is come from the faulty bearing. With the sampling rate of 32.77k Hz, the period of the impacts can be roughly estimated through dividing the time length of the signal by the number of periods in it. The result showed that impacts occurred at a time interval of around 10 ms. Based on the knowledge about the diagnosed bearing (SKF 6215) and its rotation speed (25 Hz), it has calculated that the characteristics frequency of the defected outer race is 113.6 Hz or a time interval of 8.8 ms. This value matched closely with interval of 10 ms shown by the impacts in the bottom diagram of Figure 5. Hence, the impacts are confirmed to be generated by the defective bearing. These conclusions also correspond to the experimental results that were generated by the ANC method done on the same test (see [4]). From the above results, it shows that the enhanced EVA is effective in the recovery of multiple vibration sources. Even when the vibrations generated from various mechanical components are captured as a mixture of vibration by one sensor, it is still possible to extract the diagnostically useful signals. With the help of this method, machines that have physical limitations on installing more than one sensor at a time can use such method to recover aggregated signals.

5 DISCUSSIONS AND CONCLUSIONS

It has been mentioned in the previous sections that the signal collected from the machine with one sensor is always the mixture of numerous sources, and conventional EVA cannot meets the need of recovering multiple faulty vibrations from a machine. The results of the conducted experiments show our proposed method has great potential for complex machine fault diagnosis. Based on the algorithm of our enhanced EVA, a new model of monitoring system and its schema can be proposed here. The general model of mechanical system of analysis is presented in Figure 6 with multiple sources and the data of a single sensor is in consideration. Each source signal $s_i(k)$ that represents the signal generated from a different component of machine has its own propagation channel with an additive noise. Therefore, the signal that is collected by a single sensor will be a composition of various sources and noises with different time delays. Only when all other signal components are small enough to keep the dominance level of vibration in received signal, the conventional EVA can achieve successful result. Otherwise it is difficult to recover any overwhelmed signal of interest. This is the main predicament that must be concerned before applying the conventional EVA. Our proposed method can enhance the effectiveness of conventional EVA in machine fault diagnosis through the construction of a single sensor input and multiple recovery outputs. The suggest schema is illustrated in Figure 7. Only one sensor is used to collect the mixed signals, minimizing the requirement for numbers of sensor in many practical applications. The equalizers are respectively used to eliminate the distortion that is introduced in each propagation path between the source signal $s_i(k)$ and the sensor signal $x(k)$. The recovery of multiple sources based on this proposed structure will be very useful for practical machine fault diagnosis.

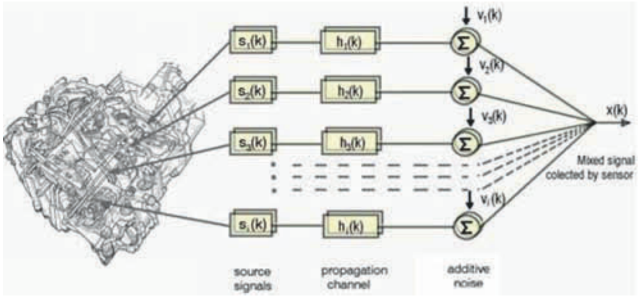


Figure 6. Multiple vibrations generation in a machine.

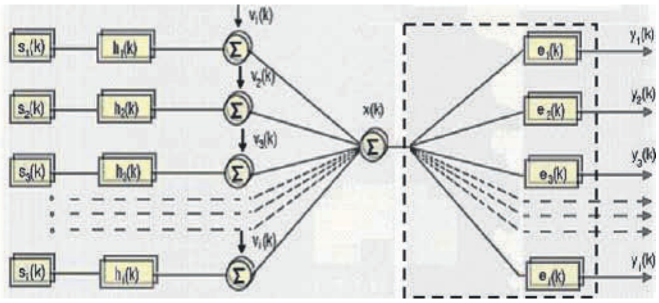


Figure 7. The schema of proposed recovery method.

Compared with the conventional EVA, the enhanced EVA in this research shows to have more prospects for complicated vibration recovery. It provides the possibility to recover more than one source signal even only using a single sensor. Such kind of multiple recoveries is impossible in conventional EVA except using Blind Source Separation with many sensors. However, there are still some problems to be overcome in its efficient application, such as the stability of signals in the presence of a high level noise, and the amplitude/scale indetermination before and after the recovery process. Further researches on EVA are required to enhance the recovery of useful vibrations for machine fault diagnosis.

6 REFERENCES

[1] Tse P., Yang W.X., & Tam H.Y. (2004) Machine fault diagnosis through an effective exact wavelet analysis. Journal of Sound and Vibration, 277, 1005-1024.

[2] Fabry P. & Serviere C. (2000) Blind separation of noisy harmonic signals using orthogonal techniques for rotating machine diagnosis. Proceedings of the Second Workshop on BSS and ICA (ICA'00), Helsinki, Finland, 369-374.

[3] Lai W. X., Tse P., Zhang G. C. & Shi T. L. (2004) Classification of gear faults using cumulants and the radial basis function network. Journal of Mechanical Systems and Signal Processing, 18, 381-389.

[4] Tse P. & Lai W. X. (2003) Minimizing interference in bearing condition monitoring and diagnosis by using adaptive noise cancellation. Proceedings of the 16th International Congress and Exhibition on Condition Monitoring and Diagnostic Engineering. Management, Vaxjo, Sweden, 739-748.

[5] Jelonnek B., Boss D. & Kammeyer K. D. (1997) Generalized eigenvector algorithm for blind equalization. Signal Processing, 61, 237-264.

[6] Jelonnek B. & Kammeyer K. D. (1994) A closed-form solution to blind equalization. Signal Processing, Special Issue on High Order Statistics, 36, 251-259.

[7] Zhang J. & Tse P. (2003) Detection of incipient impulsive fault by blind equalization. Proceeding of IMS2003, Xi'an, China, 25-27.

[8] Tan A., Yang B. S., Mathew J., Son J. D. & Kim D. J. An experimental evaluation of blind equalizer filter length for the recovery of rolling element bearing signals masked by noise, Queensland University of Technology, Australia (Paper No. 130).

[9] Tugnait J. K. (1995) On blind identifiably of multipatch channels using fractional sampling and second order cyclostationary statistic, IEEE Trans. Inform. IT- 41, 308-311.

[10] Treichler J.R., Fijalkow I. & Johnson C. R. (1996) Fractionally spaced equalizers. IEE Signal Processing Magazine, 1053-5888.

[11] Shalvi O. & Einstein E. (1990) New criteria for blind deconvolution of nonminimum phase systems (channels). IEEE Trans. Inform. Theory, IT-36, 312-321

Acknowledgments

The work that described in this paper was fully supported by a grant from the City University of Hong Kong, Hong Kong, China (project no. 7001808).

DEFECTIVE EVALUATION OF CONCRETE STRUCTURE WITH RISK-BASED MAINTENANCE

K Toda T Nishido & Y Moroyama

Ishikawajima-Harima Heavy Industries Co., Ltd, Structure & Strength Department Research Laboratory, 1 Shin-Nakahara-cho Isogo-ku, Yokohama 235-8501 Japan

Abstract: Effective maintenance for concrete structures is important and urgent tasks. Maintenance cost depends on its start time. Conventional inspection for structural defects consist of non-destructive tests, e.g. Schmidt hammer test, and destructive tests, e.g. cored samples. Defective degrees are decided by total ratings with the tests, but the tests cannot consider the importance of the structures. A concrete slab in a bridge in an urban area is more important than it in a rural area, because the destruction of the concrete slab in the urban area greatly influences on users. The size and kind of damage on the destruction are different according to the roles of the structures. The risk-based maintenance (RBM) is used for the defective evaluation of the concrete structures. If a disaster occurs, the damage can be kept minimum with the RBM. The risk evaluation is expressed a matrix, in which a vertical axial is likelihood for the destruction and a horizontal one is consequences for the damage. The conventional inspection is used only the likelihood that is classified at four levels, e.g. Safety, usefulness, influences for users and aesthetics are evaluated numerically as the consequences. By using the RBM, the effective maintenance is possible and repair cost is cheaper than with the conventional plans.

Key Words: Risk-based maintenance, Concrete structures, Defective evaluation

1 INTRODUCTION

The size and kind of damage are evaluated according to the social roles of concrete structures. The concrete structures that are located in or near sea are especially damaged by salt attack. Chloride ions from the surfaces penetrate the insides and cause steel bars to corrode. The steel bar corrosion is serious problem for reinforced concrete (RC). Once the corrosion occurs in the RC, it causes cracks and chloride ions to penetrate easily. Moreover, the heavy steel bar corrosion causes covers to come off and durability to decrease. Engineers must consider the damage and the influences on the destruction.

Risk-Based Inspection (RBI) and Risk-Based Maintenance (RBM) enable the engineers to reduce risk by considering safety and cost. The methods have been used for the maintenance of boilers, cement plants, LPG plants and other large-scale structures. The RBM is a useful tool for the operation of plants [1] especially. In recent years, administrators have considered life-cycle cost (LCC) and asset management for the concrete structures. Japan Concrete Institute (JCI) proposes the guideline of the asset management for the concrete structures [2]. The authors have applied the RBM to the concrete structures like a bridge, a shipyard dock and a plant foundation and confirmed its effectiveness in the study.

2 RBM

The RBM method has been generated to optimise inspection and maintenance plans. Risk is commonly defined as the likelihood of failure multiplied with the consequence of failure. Fuji and Kihara have suggested that if the administrators use the RBM methods for plants and other structures, cost and risk are reduced [3]. There are some approaches to the evaluate risk in the RBM (i.e. qualitative, semi-quantitative and quantitative). The fundamental differences among these approaches are the amount and the detail of input, calculations and output. Expert judgements are used to assign rating on the likelihood and the consequences of failure in the qualitative (or semi-quantitative) RBM. The expert judgements may be powerful methods and important factors to evaluate the risk, because unknown factors to assess the likelihood and the consequence of failures are included in their experiences. Figure 1 shows the procedure of the RBM approach, which may also be similar to a generic representation of a risk analysis. The steps of the procedure are summarised below [4].

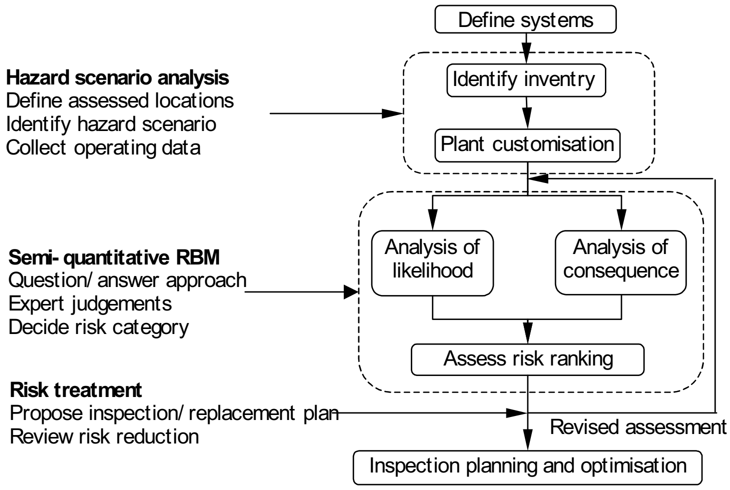
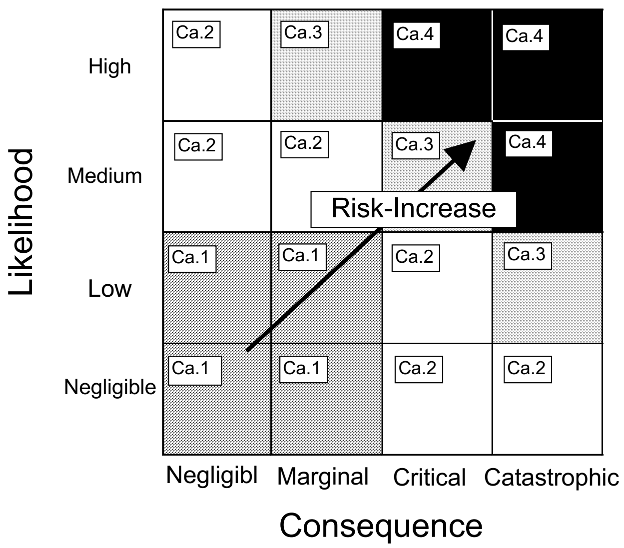


Figure 1. Flow of Risk-Based Maintenance approach



Risk Rank	Required actions
Category 1: Acceptable	No inspection or other actions required within timeframe considered unless to satisfy national legislative requirements
Category 2: Acceptable with controls	Take appropriate inspection or other actions to keep operating
Category 3: Undesirable	Mitigate to risk ranking 1 or 2 within timeframe of next inspection
Category 4: Unacceptable	Mitigate to risk ranking 1 or 2 immediately

Figure 2. Matrix of risk ranking and categories

Step 1 : Making hazard scenario analyses at assessed locations, we identify hazard scenarios and collect operating data. For the concrete structures, the worst scenarios are falling bridges and fatal or serious accidents by them.

Step 2 : The next step is question and answer approach by the expert judgements and decided the risk categories. For the decision of the likelihood degrees, quantitative and semi-quantitative data are gathered by visual and detail inspection. For the decision of the consequence degrees, we assume the risk for human and cost. Once the likelihood and the consequences are classified for a matrix as shown in Figure 2, the highest risk conditions are clear.

Step. 3 : If evaluated risk rating for a location is high (i.e. category 3 or 4 in Figure 2), the risk reduction is required. In order to reduce the risk, detail inspection and a replacement plan can be proposed. If the reduction can be confirmed, proposed methods are applied to the optimised maintenance plan. However if some plans are proposed and their revised risk ratings are same categories, the quantitative analyses may be needed to identify the optimised plan.

3 LIKELIHOOD FOR CONCRETE STRUCTURES

Assessment is needed to classify the damage on the likelihood for the concrete structures. The first step is to check design books, which include situations, locations, construction years, design strength, water-cement ratios, sorts of aggregates and so on. The information in the design books is useful for the classification. The second step is inspection for the concrete structures. Diagnosis items are shown at Table 1. The classification is mainly done by the visual inspection [5]. Although non-destructive tests contain unknown factors, they are often useful. If cracks found by the visual inspection on the concrete structures have large width (0.2mm), chloride ions and carbon dioxide can easily cause steel bar corrosion. If the compressive strength measured by the Schmidt hammer test is low, the likelihood is judged as high.

Destructive tests are done for the detail inspection. If sampling time and place are permitted, cored simple tests, which indicate objective compressive strength and Young' module, is the best method for the diagnosis. The cored sample tests are also needed repairs because the tests damage the concrete structures.

Table 1 Diagnosis tems

Classifications	Items	Methods
Non-destructive tests	Inspection for degree of damaged	Visual Inspection
	Compressive strength	Schmidt hammer test
	Crack depth, Space between concrete and steel bar	Ultrasonic method
	Carbonate degree at exposure point	Phenolphthalein method
Destructive test (Cored sample)	Compressive strength	Compressive strength test
	Carbonate depth	Phenolphthalein method
	Amount of chloride ion	Potential difference titration
	Alkaline aggregate reaction	Microscope inspection

Table 2 Damage states and grades of visual inspection for salt attack and carbonate

Damaged grades	Salt attack	Carbonate
Negligible	Damage is not detected by visual inspection. Amount of chloride ions is less than corrosive occurrence amount.	Damage is not detected by visual inspection. Depth of carbonate is below cover.
Low	Damage is not detected by visual inspection. Amount of chloride ions is more than corrosion occurrence amount.	Some cracks and rust water can be seen. Depth of carbonate is near cover.
Medium	Some cracks occur by steel bar corrosion. Rust water can be seen.	Many cracks and rust water can be seen. Depth of carbonate is near steel bars.
High	Many and wide cracks occur. Loss of covers can be seen.	Many and wide cracks occur. Loss of covers can be seen. Depth of carbonate is over steel bars.

4 CONSEQENCES FOR CONCRETE STRUCTURES

The consequences differ on situations, the types of the concrete structures and damage mechanisms, so the definitions of the consequences are important for the RBM. There are three examples, a bridge, a shipyard dock and a LPG plant foundation.

4.1 Bridges

Concrete slabs in bridges are heavily damaged by fatigue, water and chloride ions in marine environment and with snow melting medicine. Once cracks occur at the surfaces of the slabs, corrosion factors like water and salt penetrate the concrete and cause steel bars to corrode. Stewart has studies chloride-induced corrosion models for bridges [6]. The risk on the consequences is considered as following. The items are objectively evaluated.

a) The loss of covers by the steel bar corrosion and disaster for users.

b) The cost of repair

c) The amount of traffic

d) The importance of road (national or local road)

e) The existence of another bridge under the bridge

4.2 Shipyard Docks

Shipyard docks are often damaged by the salt-attack, because of dry and wet repeat on sea conditions. Cracks can easily occur by dry shrinkage, even though the covers is thick. Then chloride ions can diffuse easily the cracks. If cement-water ratio is more than 50%, the cracks occur regardless of the steel bar corrosion. The risks on consequences are considered as following.

a) The loss of covers and the disaster for workers at shipyard dock bottoms

b) The cost of repair increases because areas are wide and high generally. If repair locations are high, scaffolds or ladder trucks are needed.

4.3 Plant Foundations

Concrete foundations at plants are constructed based on the designed compressive strength of the concrete. If actual compressive strength is lower than designed one, wide and deep cracks occur on their surfaces. If thermal stress is over allowable tensile stress, wide and deep cracks also appear on their surfaces. Chloride ions, carbon dioxide and other corrosion factors diffuse the concrete insides thorough the cracks. Therefore low compressive strength and excessive thermal stress weaken their durability. Plants have many equipments on their concrete foundations, so the risk on consequences as following [7].

a) The amount of flammable solution

b) Heights

c) Boiling solution

d) Driving pressure

e) Rotation numbers

5 APPLICATION TO RBM FOR CONCRETE WALLS IN SHIPYARD DOCK

The authors have investigated the large concrete walls of a shipyard dock that was built in 1972. It has three concrete walls and is stored seawater by a steel gate. The concrete walls are repeated on dry and wet conditions for almost 30 years. A large amount of chloride ions diffuses the concrete wall insides. The designed cover thickness is over 50mm, but serious steel bar corrosion occurs and the covers come off at some places. Repair and replacement have been done according to the damage without thinking of the risk for the repair and workers. The authors apply the RBM to reduce the risk and decide the orders of the repair on the walls.

5.1 Design Books and Document Data

The authors have checked the design books and the document data on the dock. The results are as following.

a) Construction year: 1972

b) Cement type: Portland-blast furnace cement

c) Design cover thickness: Over 50mm

d) Type of aggregate: Blast-furnace slag coarse aggregate

e) Water-cement ratio: 62.5%

f) Design compressive strength: 24N/mm2

5.2 Visual Inspection for Likelihood

The northern and the southern walls are divided in 4×25=100 and 4×22=88 blocks, and the eastern wall in 5×4=20 blocks respectively to reduce operations for diagnosis.

The visual inspection is the best and simple method to investigate the degree of damage considering the size and the area of the dock. The damaged states and grades with the visual inspection are evaluated by using JCI (Japan Concrete Institute). There are many cracks and loss of covers by the steel bar corrosion (Figure 3). The square blocks are ranked by using Table 2.

a) Medium ranked stage

b) High ranked stage

Figure 3. Damage by salt attack.

5.3 Risk for Workers at Bottoms

If the loss of covers by the steel bar corrosion, The workers received serious damages (injured or dead). The risk on the consequences for workers is

$$R_{tp}= S_d \times H_d \times P_{tp} \quad (1)$$

where R_{tp} is the risk for workers, S_d is the damage of one concrete block, H_d is the height of damaged area and P_{tp} is the number of the workers in a day. If the workers at the bottom are near stairs, the risk for them increases. All the blocks near the stairs are ranked as four.

5.4 Risk for Repair

The area of the dock is so large. If all the blocks are repaired, the cost increases. The RBM may be valuable for the economical maintenance of the dock. Main repair for the area, in where there are the loss of covers and spaces among the concrete and the steel bars, is to replace damaged concrete with new one. The risk on the consequences for the repair is

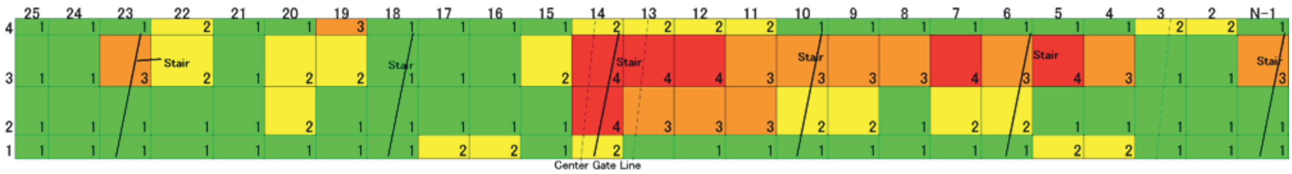
$$R_{rp}=S_d \times C_{rp}+H_d \times C_{sc} \quad (2)$$

where R_{rp} is the risk of the repair, C_{rp} is the cost of the repair and C_{sc} is the cost of the scaffolds or ladder trucks. The cost of the repair is containing materials and persons. It is very important for administrators to choose proper repair methods. If partial repair for the steel bar surfaces is needed, corrosion protect admixtures like nitrite on the surfaces is better. Damage will occur again due to careless repair. Therefore the cost may rise up than the evaluated one.

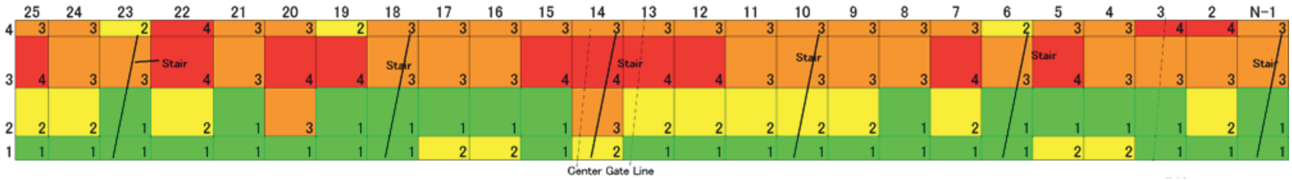
5.5 Results of RBM

Figure 4 shows the results of the RBM at the southern and the northern walls. Numbers are calculated by using the likelihood and the consequences. The numbers mean 1: acceptable, 2: acceptable with controls, 3: undesirable and 4: unacceptable, respectively. The numbers also show the orders of priority for the repair on the walls. The blocks near the stairs and the heights are evaluated high ratings considering danger for the workers. Figure 5 shows the summarised risk matrixes of the RBM. The ways to reduce the risk ratings of the matrixes is partially repaired the blocks that are heavily damaged.

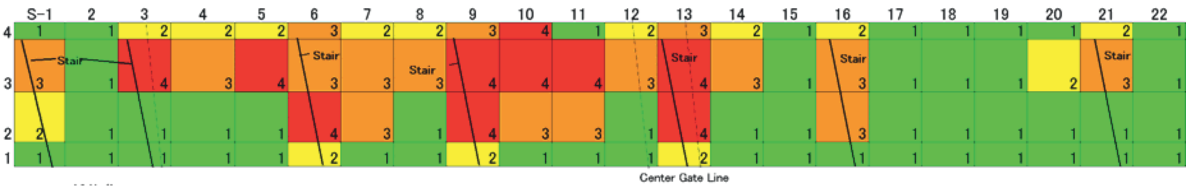
Desalinate is one of better electrochemical repair methods, but it needs large area. Therefore it must be only used for the highest risk ratings. The ratings can be reduced to keep the workers out the stairs during repairing or repair first at the heavily damaged blocks. The results of the RBM may be helpful for administrators to properly choice the ways to reduce the risk.



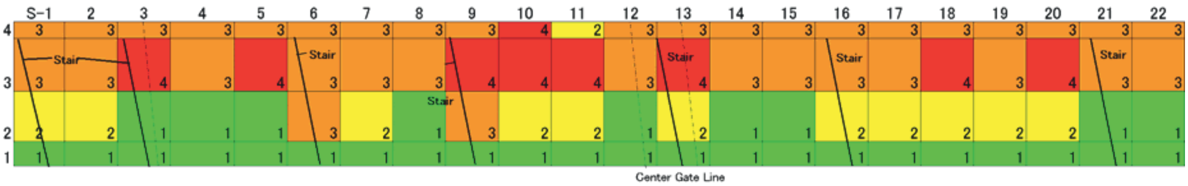
a) Risk for workers at northern wall bottom



b) Risk for repair at northern wall

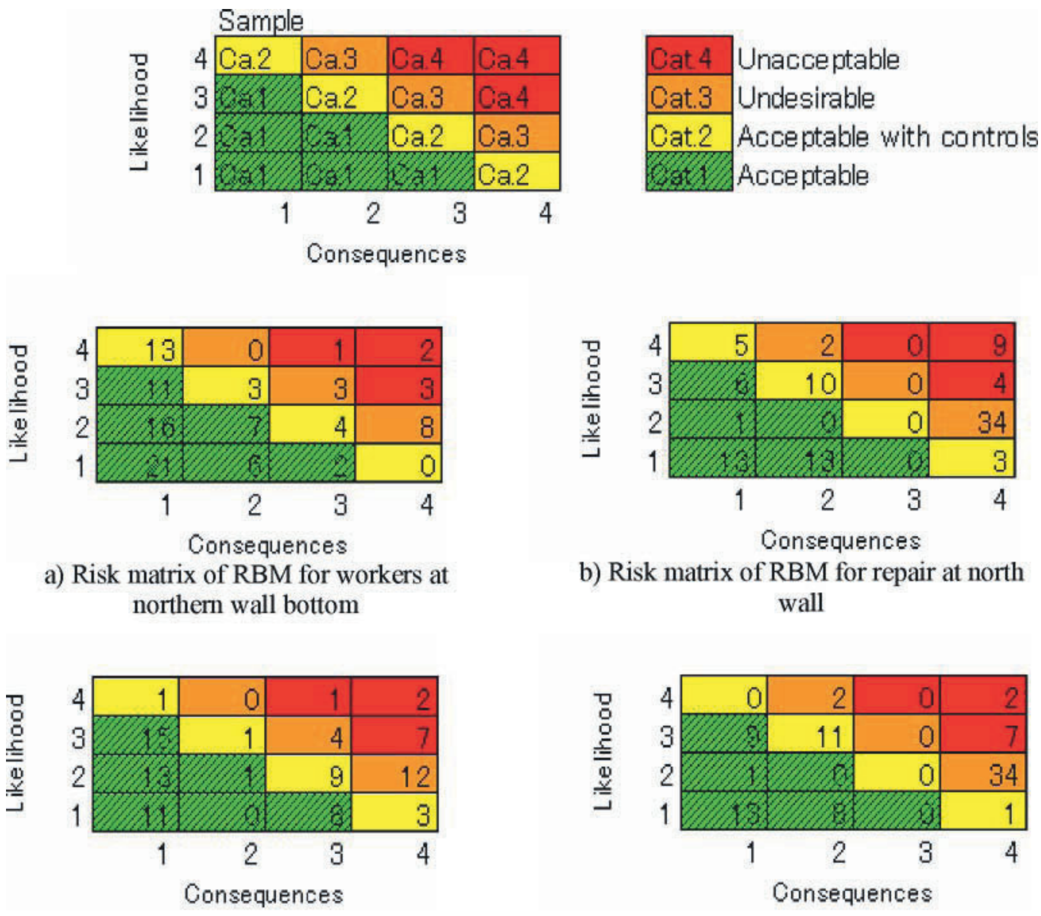


c) Risk for workers at southern wall bottom



d) Risl for repair at souther wall

Figure 4. Results of RBM application for northern and southern shipyard dock walls



a) Risk matrix of RBM for workers at northern wall bottom

b) Risk matrix of RBM for repair at north wall

c) Risk matrix of RBM for workers at southern wall bottom

d) Risk matrix of RBM for repair at southern wall

Figure 5. Risk matrixes of RBM for shipyard dock.

6 APPLICATION TO RBM FOR CONCRETE STRUCTURES IN BRIDGES

The damage of bridges influences not only repair cost but also economical activities. Therefore the bridges are one of the most important concrete structures. If inspection is carried out, traffic regulations may be needed and the flows of cars may go bad. Administrators must avoid the risk during the inspection. Non-destructive tests are useful methods [8] for the diagnosis of the concrete structures compared to destructive tests like cored samples because traffic regulations are minimum.

7 NON-DESTRUCTIVE TESTS FOR CONCRETE STRUCTURES IN BRIDGES

Non-destructive tests for the concrete have been improved and have various methods. Places and methods are shown in Figure 6 when they are applied to a bridge.

7.1 Detail Inspection for Bridges

If chloride ions penetrate the concrete insides and till steel bar surfaces and the amount of the chloride ions more than 1.2kg/m^3 on the surfaces, the steel bars corrode. Cracks due to the corrosion occur; the chloride ions penetrate from there easily. The authors have investigated a steel bridge with a concrete slab, which was constructed 40 years ago. The bridge is situated at a mountain area, so calcium chloride has been sprinkled for melting ice and snow on the slab. The chloride ions in the curb on the slab are shown in Table 3.

Table 3 Amount of chloride ions and calculated-diffusion coefficient.

Sampling position from surface	Mass per unit volume	Amount of chloride ions	Diffusion coefficient of chloride ion	Chloride ion concentration on surface
(mm)	(kg/m^3)	(kg/m^3)	(cm^2/year)	(kg/m^3)
10 – 20	2254	4.40	1.452	5.40
30 – 40		2.46		
50 - 60		2.05		

The diffusion of the chloride ions in concrete is said to follow Fick's second diffusion equation, as expressed with Eq.(3)

$$C(x,t) = C_0\left(1 - erf\,\frac{x}{2\sqrt{D \cdot t}}\right) \tag{3}$$

where C(x,t) is the concentration of the chloride ion (kg/m^3) at depth x (cm) and in period t (year), C_0 is the chloride ion concentration (kg/m^3) on surface, erf is the error coefficient and D is the diffusion coefficient of the chloride ion (cm^2/year). If cover thickness from a steel bars is known, the amount of the chloride ions can be calculated by Eq.(3). The start year of corrosion and the life of a bridge are predicted.

The cover of thickness in the curb, which was reconstructed 10 years ago, is 60mm. The maximum crack width on the surface is 0.5mm. Calculated chloride ion by Eq.(3) on the steel bars is $1.9kg/m^3$, so corrosion started about 4.7 years ago. The heavily corrosion cannot be seen at the surfaces of the steel bars actually, but corrosion will continue and let the cracks to occur from the locations where the chloride ions are calculated.

7.2 Diagnosis for Bridges

If the results are needed high accuracy, Schmidt hammer and cored sample tests are useful for the quantitative. Specialists decide the degrees of the likelihood with their results. The consequences differ at situations where bridges are located as explained in 4.1. The RBM method is applied to various bridges for its accuracy and reliability, as the results of the RBM are omitted here.

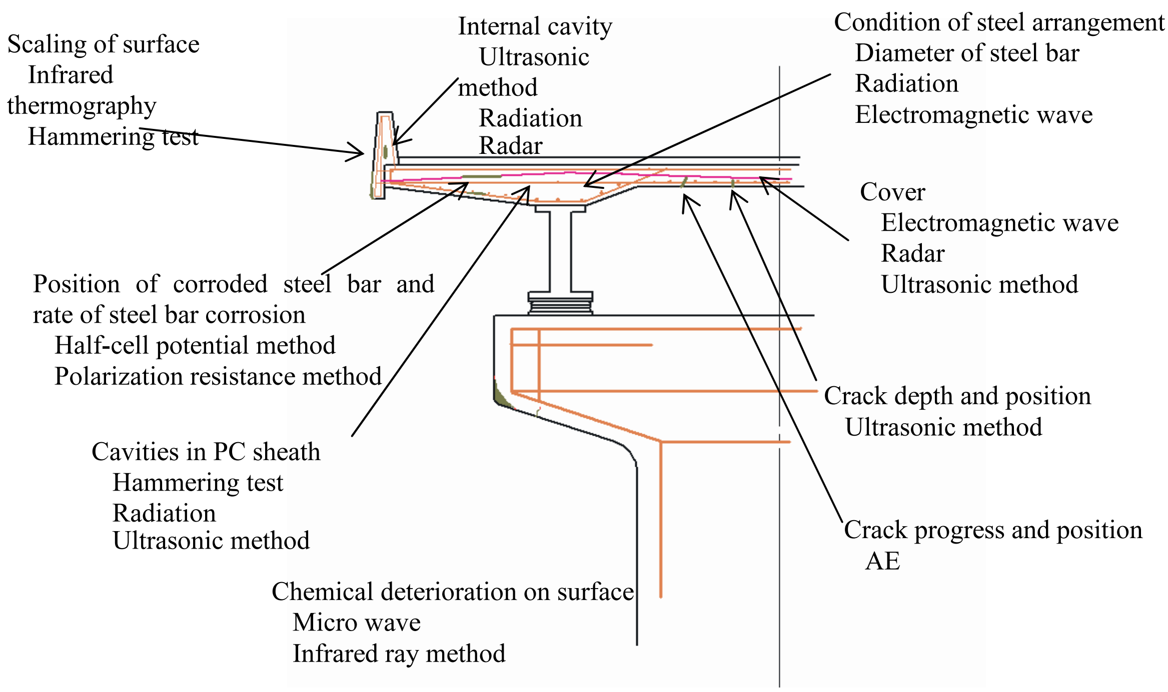


Figure 6. Non-destructive tests for bridge

8 CONCLUSIONS

Although the concrete structures have been said maintenance free, there are many damaged concrete structures. JCI has suggested the standard judgements of diagnosis for defective structures. If the RBM method is used for the concrete structures, the highest risk ratings are clearly assessed by using the risk matrixes. Therefore administrators can propose where is first selected and repaired. Namely they can decide some repair methods to reduce the risk rankings at the shipyard dock.

The detail inspection by the cored samples is the best method for the diagnosis, but cost, time and fixing the cores are needed. Speedy inspection is needed for the bridges to reduce traffic regulations. The diagnosis may be done only by the visual inspection. As the consequences differ at the locations of the bridges and are especially complicated, a guideline for them must be made. Further study on the RBM is needed for more accurate defective evaluation of the concrete structures.

9 REFERENCES

[1] S. Kihara and A. Fuji (2002) A guide to RMI/RBM, maintenance by risk assessment, Japan institute of plant maintenance.

[2] Japan Concrete Institute session (2005) Research committee report of risk management for concrete structures.

[3] A. Fuji, H. Eguchi, C. Fukuoka, T. Tanoue, J. Takahashi, Y. Okatsuka and S. Kihara (2001) Application of risk based maintenance with life and financial assessment on fossil-fired power plants, Advance in fracture research proceeding of ICF 10.

[4] M. Yatomi, A. Fuji, N. Saito and T. Yoshida (2004) Risk and reliability and evaluation of components and machinery, PVP-Vol.488, 47-51.

[5] Japan Concrete Institute (2002) Concrete diagnosis technique (Foundation), 181-195.

[6] Kim A. T. Vu and M. G. Stewart (2000) Structural reliability of concrete bridges including improved chloride-induced corrosion models, Structural Safety, 22, 313-333.

[7] K. Toda, J. Takahashi and A. Fuji (2005) Proposal of concrete diagnosis technique that uses RBM, Japan society of civil engineers, Annual committee Vol.60, 65-66.

[8] K. Toda, H. Hatanaka, Y. Hayashi, Y. Kawano and M. Iketani (2004) Diagnosis of concrete deterioration by non-destructive evaluation, IHI Engineering Review, Vol.37, No.3, 98-103.

DEVELOPMENT OF FIBER OPTIC INCLINOMETERS FOR BRIDGE MONITORING

Eiichi SASAKI[a] , Ryo TANAHASHI [a], Yuji ISHIKAWA [b], Sanae MIYAZAKI [b], and Hitoshi YAMADA [a]

[a] *Dept of Civil Engineering, Yokohama National University, 79-7 Tokiwadai, Hodogaya-ku, Yokohama, 240-8501, Japan.*

[b] *Research and Development Headquarters, NTT DATA CORPORATION, Kayabacho Tower Bldg., 21-2 Shinkawa 1 – chome, Chuo-ku, Tokyo 104-oo33, Japan.*

Abstract: It is important to develop an efficient maintenance strategy for existing infrastructures. In this study, a new sensor was developed to establish a new concept of health monitoring of existing bridges. Bridges always show 3 dimensional dynamic behaviours with deflection or torsion by external turbulences such as traffic, winds, and earthquakes. Temperature is also an important factor that governs bridge motions. In order to detect and analyse damage and conditions of bridges, it is essential to monitor these 3 dimensional dynamic responses of bridges. Especially, dynamic torsional behaviours of bridges have not been discussed on so far, but it is expected that they give much information on damage and conditions of bridges. In this study, a new fiber optic sensor, that is, an fiber optic inclinometer was proposed for a new bridge monitoring strategy based on torsional behaviours. Through the tests on dynamic characteristics and temperature effects, the new sensor was developed and verified.

Key Words: inclinometers, bridge, monitoring

1 INTRODUCTION

Lately, the demand for the maintenance of existing infrastructures has been increasing all over the world. It is about time most of existing infrastructures need to be maintained. However, it needs enormous costs and time to maintain so many infrastructures. Therefore, it is important to develop an efficient maintenance strategy for existing infrastructures.

In Japan, the fiber-optic communications net has been connected along the national highways to maintain infrastructures. It will be extended to 300,000 km by 2010. Miki et el (1) monitored an instrumented bridge with the fiber-optic communications net (2001). They constructed the remote monitoring system that can forward the real-time data to the monitoring station. Kobayashi et el (2) attempted to develop a Weigh- In-Motion system utilizing Fiber Bragg Grating (FBG) sensor (2003). Because FBG sensors have capability of dynamic measurements and high durability, it can be considered as promising sensors suitable for bridge monitoring. Recently, new vibration sensors using FBG sensors have been proposed by Mita et el (4) (2003). Furthermore, Miki et el (3) observed a bridge with a monitoring system all by fiber optic system utilizing FBG sensors (2005). These intensive surveys gave the results that the "fiber-optic" monitoring system established by the combination of FBG sensors and the fiber-optic communications network could realize long-term maintenance-free. It should be emphasized that the fiber-optic monitoring system doesn't need to set up power supply in the field, so it can avoid troubles with thunderbolts, electromagnetic noises and dust.

The already-proposed monitoring systems need many sensors to grasp behavior of bridges, because sensors basically measure strain at a point in the target. This makes the cost of the monitoring system considerable. Therefore, in order to expand the applicability of monitoring systems by reduction of the number of sensors, a new concept of bridge monitoring must be established. In this study, focusing on 3 dimensional dynamic responses, fiber optic inclinometers are proposed as a new sensor. The proposed sensors by the authors can monitor 3 dimensional behaviors of bridges, especially torsional behaviors of bridges. It can be expected that 3 dimensional behaviors can give more information to find deterioration of existing bridges than strain gauge and that. the number of sensors will reduce compared with the systems using strain gauges.

With these sensors, measurements at few points make it possible to detect the dynamic responses because of the performance of inclinometer. And the sensor can be set up easily without scaffold. And this system has multiplexing capability. It can monitor plural bridges remotely at once by a set of the system. These factors would realize cost reduction. When it comes to the accuracy, the positions to set up sensors can be changed easily at any time it is needed because of the flexibility in placing.

In this study, the dynamic response characteristics of the proposed sensor were investigated to evaluate application of the Optical Fiber Inclinometer to monitor existing bridges. In addition, because fiber optic is extremely influenced by temperature, temperature compensation of the sensor was inspected.

2 DEVELOPED INCLINOMETERS

Fiber Bragg Grating (FBG) is adopted for this sensor because of the characteristic like durability, simplicity and multiplexing capability (Moreover it doesn't need power supply in the field.) (5).

In this sensor (Showed on Fig.1), the shift of the Bragg wavelengths indicates inclination.

The mechanism of the inclinometer is showed on Fig 2. The swing arm swings as to the pendulum movement, then a strain is happened to the FBG sensor by a stress at the plate string. The direction of pendulum swing is measurable direction of the sensor.

Figure 3. displays the relation between Bragg wavelength shifts and inclination of this sensor. The transform coefficient from wavelength [nm] to inclination [degree] in certain temperature (22.3 degrees of Celsius) is 2.674 [degrees/nm]. The relation can be showed this linear equation below.

Inclination[degree]
= 2.674 [degree/nm] ×wavelength shift [nm]

Figure 1 FBG Inclinometer

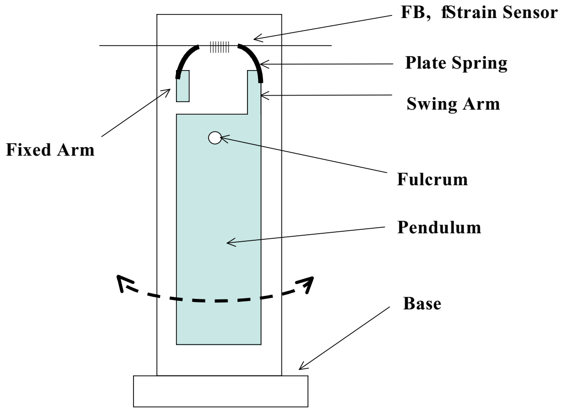


Figure 2. FBG Inclinometer Structure @

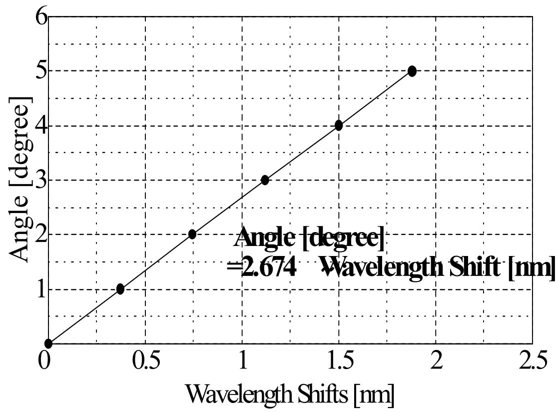


Figure 3 Wavelength Shifts and Inclination

3 TEMPERATURE CHARACTERISTICS

Generally, FBG is extremely influenced by temperature. The error by temperature makes virtual inclination. Therefore, we need to inspect the effect of temperature on the measurement to compensate for temperature.

In this test, the sensor was put into the high temperature chamber. And the temperature was changed by 10 degree of Celsius [ ] from –14 to 64 . Wavelengths were measured on each temperature.

From the results of wavelength on each temperature without inclination, Fig.4 shows relative wavelengths (wavelength shifts) to 1534.642 nm which is at 22.3 as a standard. Wavelength lengthens almost in proportion to temperature. From the result (Fig.4) and the relation between wavelength shifts and inclination (Fig.3), it can be read off that the error is from –0.14 degree to +0.28 degree between –14 and 60 (Fig.5). These wavelength shifts made virtual inclination.

4 DYNAMIC RESPONSE CHARACTERISTICS

The dynamic response characteristic of this sensor has to be investigated to grasp dynamic torsional behaviors of actual bridges. To grasp dynamic characteristics of bridges gives some information like deterioration in a girder and bearing etc. We focused on the response of inclination vibration characteristic but we also detected vertical vibration characteristic, because the sensor is expected to distinguish inclination vibration from other vibration.

Therefore, a vibration simulator was made to simulate torsional vibration of bridges (Fig.6). Two linear actuators were used for the vibration simulator. It can simulate torsional behaviour with control these actuators out of phase. And it can also simulate vertical vibration in phase with each actuator.

4.1 TEST CONDITIONS

The new software was developed to monitor the sensor's responses. Figure 7 shows the system of the dynamic responses characteristic tests.

In these tests, frequency of the vibration table was changed from 2Hz to 5Hz on each behaviors (vertical vibration and inclination vibration) according to actual bridges characteristic (5).

In the vertical vibration tests, the amplitude of acceleration of the table was controlled around 0.1[G].

In the inclination vibration tests, the angular acceleration of the table was compared with the response of FBG inclinometer.

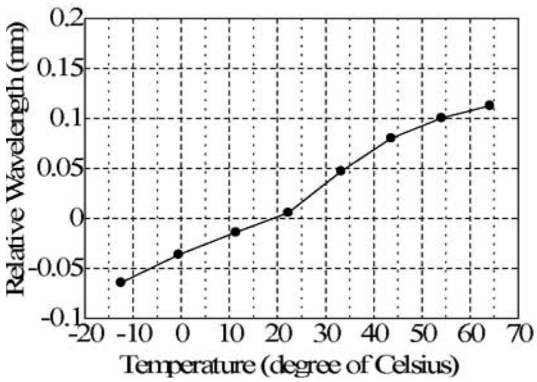


Figure 4 Temperature and Relative Wavelength

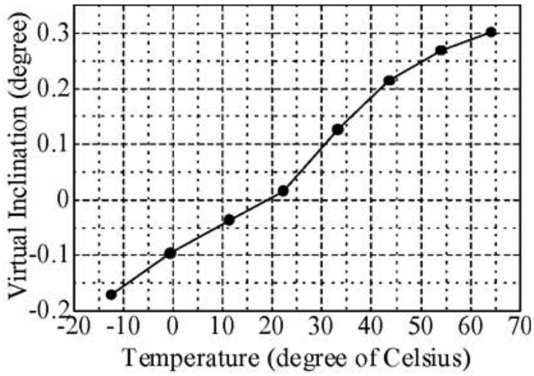


Figure 5 Temperature and Virtual Inclination

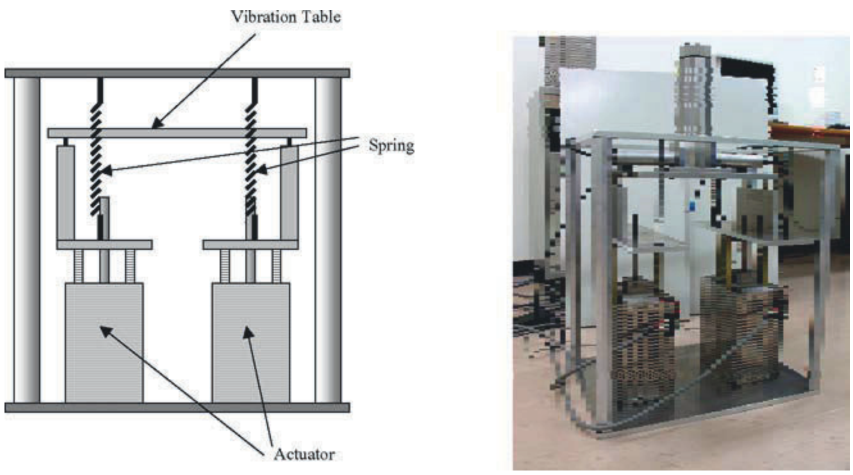


Figure 6 Vibration Simulator

4.2 TEST RESULTS

Figure 8. to Figure 13. show results of tests on dynamic response characteristics. The response of the FBG inclinometer in 2Hz and 3Hz vertical vibration (Fig.8) was extremely small compared with the case in inclination vibration (Fig.11). In addition, the result of FFT analysis in 2Hz vertical vibration (Fig.9) indicates that the small vibration seemed to come from the FBG inclinometer's own characteristic frequency (6Hz). This sensor is expected to indicate minute responses in vertical vibration. On the other hand, this sensor responded bigger in more rapid vertical vibration like 4Hz and 5Hz (Fig.8). From the results of FFT analysis (Fig.9), it can be found that the sensor grasped this table's vertical vibration frequency in 4Hz and 5Hz. Although this sensor responds in vertical vibration that is not measurable direction because the pendulum of this sensor (Fig.2) is not properly symmetry, it is possible to separate kinds of vibration by FFT analysis.

By comparison, in inclination vibration tests, instability of the vibration table's motion reflects the results especially in 2Hz and 3Hz, but from the results of FFT analysis (Fig.12 and 13) and the number of waves in Fig.10 and 11, FBG inclinometer's responses apparently corresponded to the frequency of the vibration table motion. Beside the results of frequency, compared with each angular acceleration measured by accelerometer, each inclination measured by FBG inclinometer also corresponded to actual inclination to some extent.

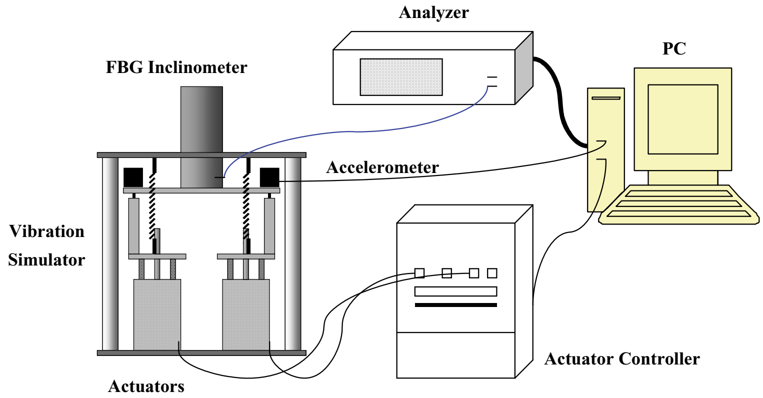


Figure 7 System of dynamic responses characteristic tests

5 CONCLUSIONS

This paper proposed the FBG Inclinometer to establish a new concept of health monitoring system for existing infrastructures. Therefore, the temperature characteristic and dynamic responses of this sensor were investigated to evaluate application of the sensor. From the results of these tests, the following facts were appeared as the characteristic of the sensor.

a) Temperature characteristic of this sensor is linear relation to wavelength shift. It is possible to compensate temperature influence of this sensor.

b) In vertical vibration, this sensor responds small in slow vibration like 2Hz and 3Hz as expectation. Although this sensor grasps vibration frequency in rapid vibration like 4Hz and 5Hz, these waves can be classified with FFT analysis.
c) In inclination vibration, this sensor grasps vibration frequency well.

The FBG Inclinometer was developed and is going to be applied to monitor an actual bridge next.

6 REFERENCES

[1] Chitoshi MIKI, Toshio MIZUNOUE and Yusuke KOBAYASHI (2001) Monitoring system of bridge performance with fiber-optic communications net, Journal of Construction Management and Engineering, JSCE, 686/VI–52, 31-40.

[2] Yusuke KOBAYASHI, Chitoshi MIKI and Eiichi SASAKI(2003) Development of Weigh-In-Motion System Utilizing FBG sensor, Journal of Applied Mechanics, 6, 1009-1016.

[3] Chitoshi MIKI, Keigo SUZUKI, Takashi KANO, Eiichi SASAKI, Minoru ISHIDA and Hiroyuki TAKAMORI: Preventive works for fatigue damage in orthotropic steel bridge deck by SFRC pavement and long term monitoring of the composite action, Journal of Construction Management and Engineering (submitted).

[4] Hiroki KAZAMA and Akira MITA (2003) Micro vibration sensor system using Fiber Bragg Grating, Journal of Applied Mechanics, JSCE, 6, 993-1000

[5] Yuji ISHIKAWA, Eiichi SASAKI, Sanae MIYAZAKI and Hitoshi YAMADA (2006) Estimation of bridge vibration using fiber optic inclinometers towards bridge health monitoring, (submitted)

[6] Hiroshi HAYANO and Akira MITA (2003) FBG Strain Sensor with Temperature Compensation, Journal of Applied Mechanics, JSCE, 6, 1001-1008.

[7] Hideyo SUZUKI and Maria Q.FENG (1995) Development of Optical Fiber Sensor for Infrastructures using Vibrating Wire and Electric Resonance Circuit, Journal of Construction Management and Engineering, JSCE, 525/VI-29, 7-15.

[8] Subcommittee on Investigation of Bridge Vibration Monitoring: Research Trends on Intelligent Bridge/Structure and Smart Monitoring (2000), Journal of Structural Mechanics and Earthquake Engineering, JSCE, 654/I-52, 1-12.

[9] Tatsuro KOSAKA (2003) Fiber Optic Strain Monitoring of FW Composite, Journal of The Japanese Society for Non-Destructive Inspection, 54/4, 182-187.

[10] Hiroshi TSUDA : Development of Fiber Bragg Grating Sensor for Structural Health Monitoring, Journal of The Japanese Society for Non-Destructive Inspection, Vol.54/2, pp.71-75, 2005.2.

[11] Hiroshi TSUDA, Junji TAKATSUBO, Nobuyuki TOYAMA and Kei URABE: Response of FBG Sensors to Lamb Wave Propagation and its Application to Active Sensing, Journal of The Japanese Society for Non-Destructive Inspection, Vol.53/7, pp.419-425, 2004.7.

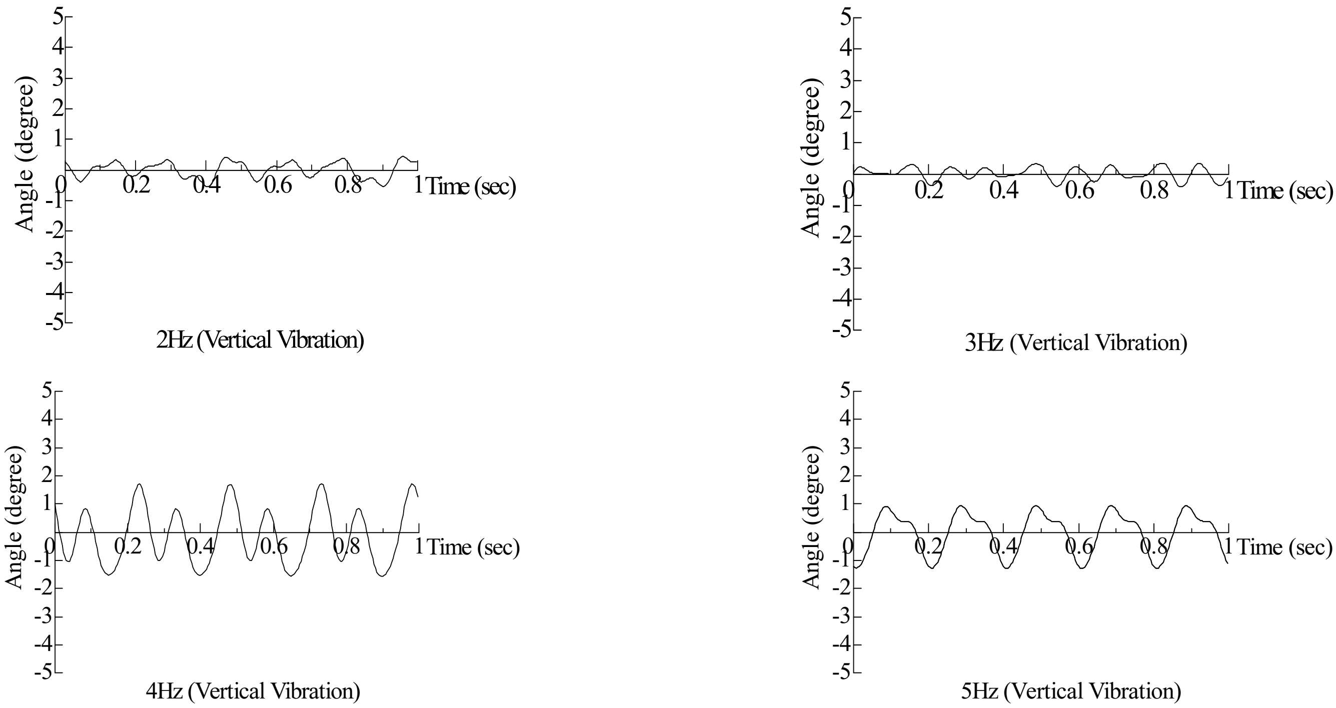


Figure 8 Response of FBG Sensor in Vertical Vibration (2Hz~5hz)

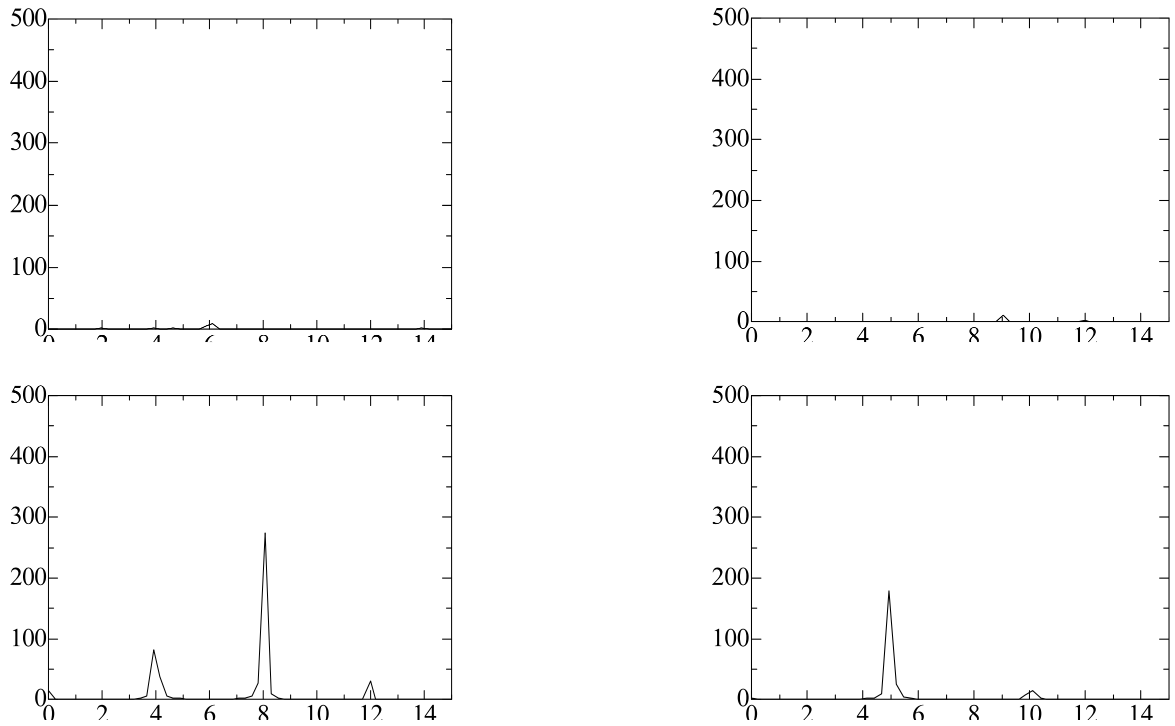


Figure 9. FFT Analysis [FBG Inclinometer in Vertical Vibration]

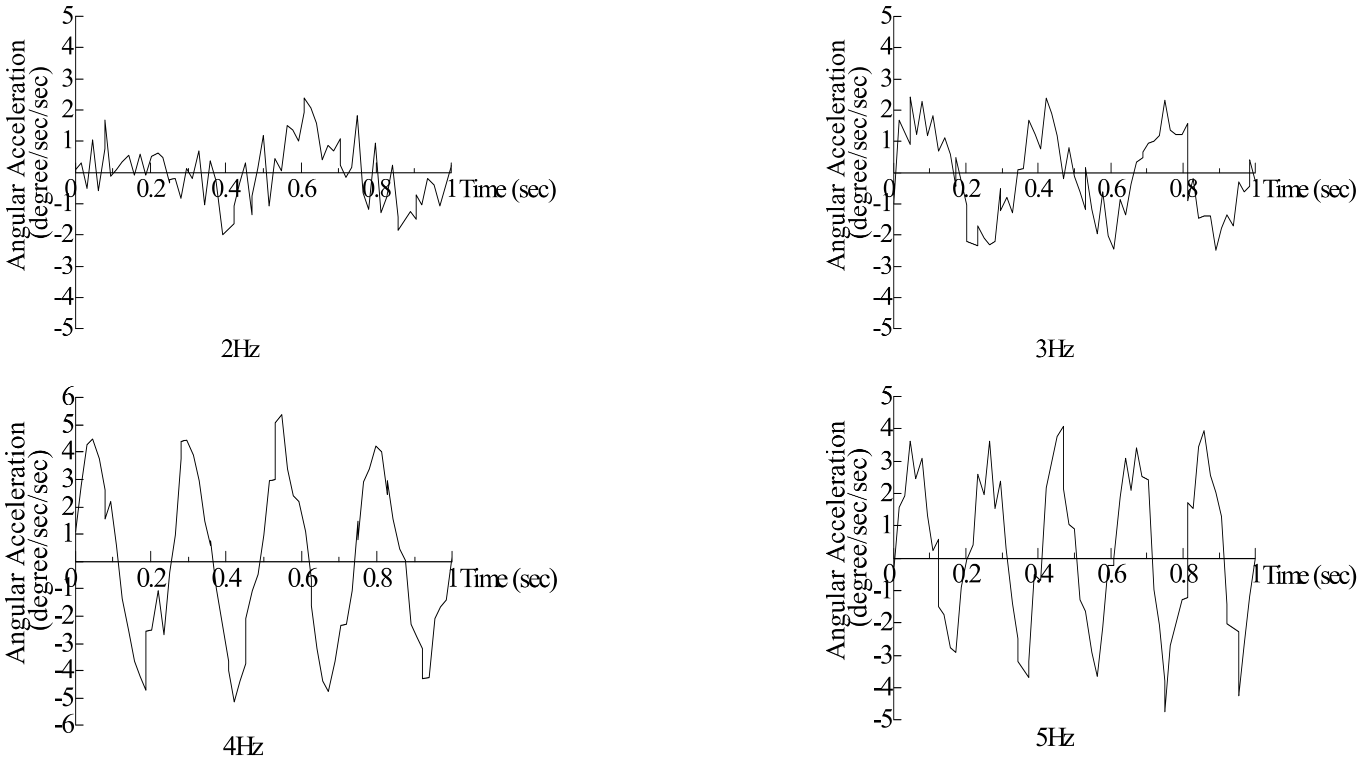


Figure 10 Response of the Accelerometer in Inclination Vibration (2Hz~5Hz)

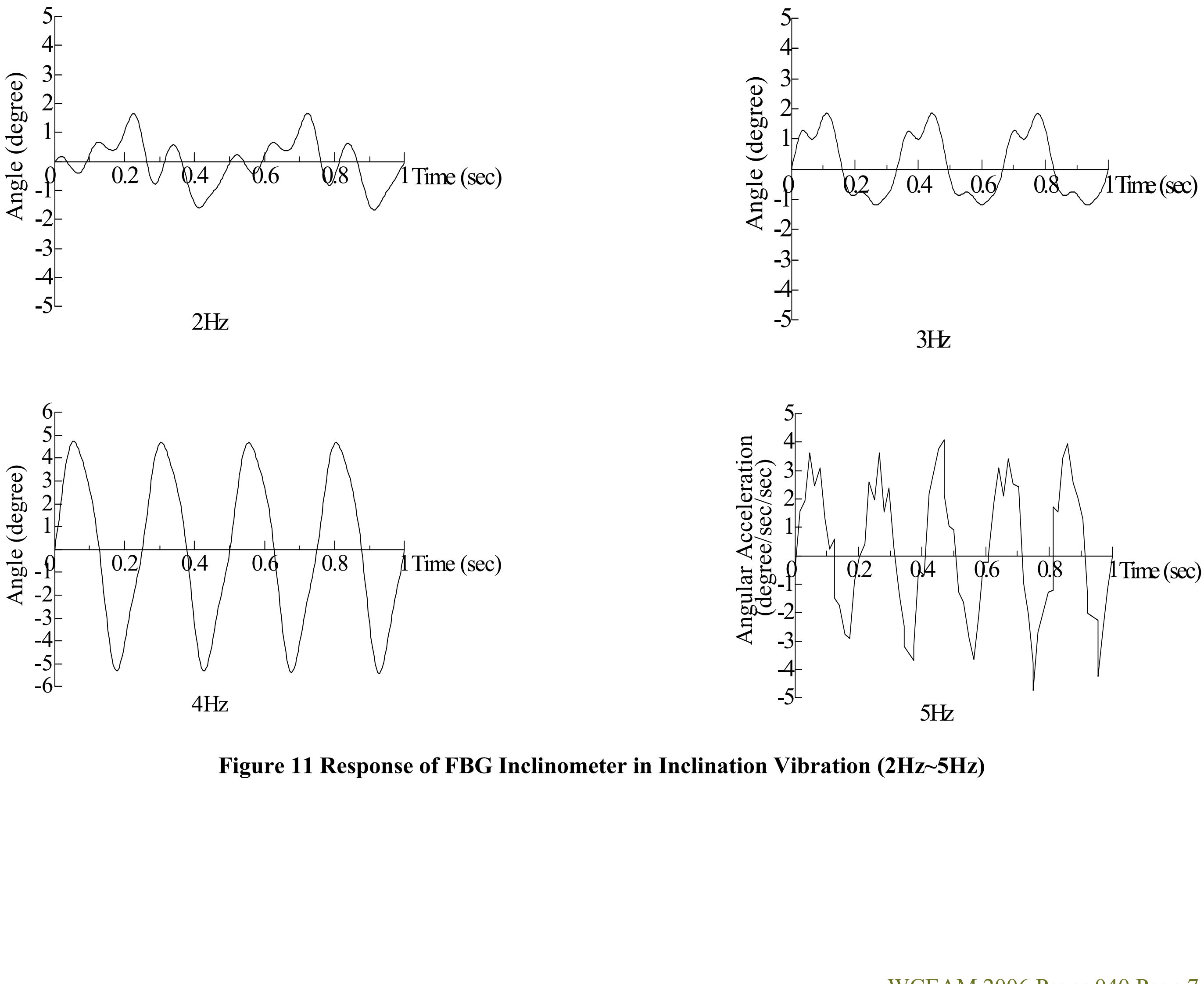


Figure 11 Response of FBG Inclinometer in Inclination Vibration (2Hz~5Hz)

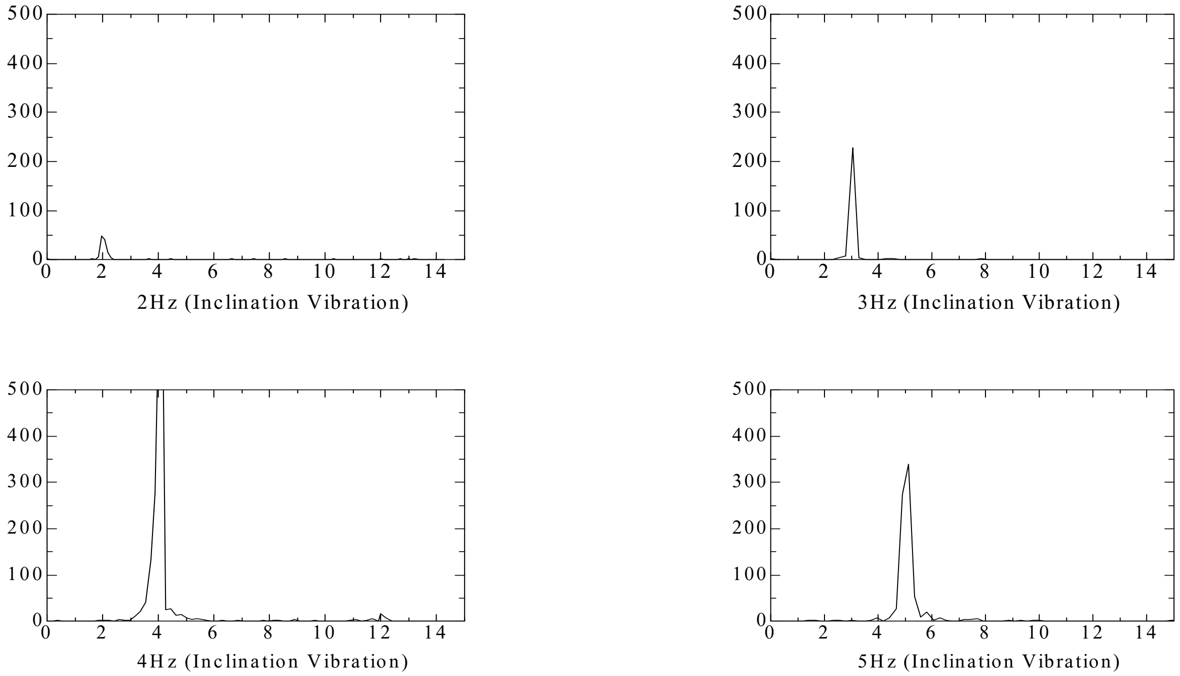


Figure 12. FFT Analysis [Angular Acceleration in Inclination Vibration]

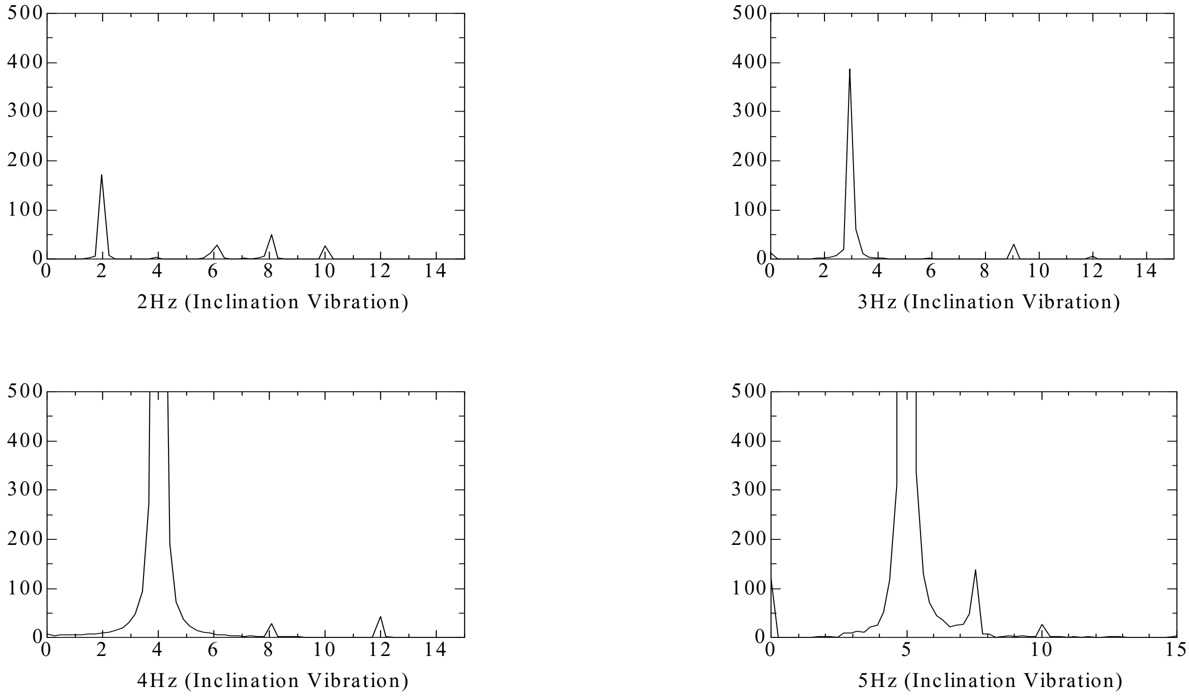


Figure 13. FFT Analysis [FBG Inclinometer in Inclination Vibration]

OPTIMISING ASSET MANAGEMENT DECISION MAKING AND BUDGETING USING RISK MANAGEMENT TECHNIQUES

Adam Aspinall and Peter Trueman

Evans & Peck Pty Ltd, Suite 5, 165 Moggill Road, Taringa, Queensland 4068, Australia

Abstract: The management, maintenance and operation of major infrastructure assets involves long (whole-of-life), complex processes in which there is always a considerable uncertainty for the manager of those assets. The level of uncertainty of asset performance, cost of operation and market competitiveness depends on risk management measures that are put in place (whether or not they are recognised as risk management measures). However, it is rare that risk management measures fully remove these risks. For a long time now, sophisticated levels of quantitative risk analysis have been adopted as tools to both improve management and make financial decisions in other business realms, particularly in the finance and insurance industries and more recently in major engineering capital works projects with considerable success. Quantitative residual risk analysis is now considerably more accessible in the broader engineering context as a result of readily available, easy-to-use computer programs and a growing body of knowledge of the use of modelling techniques. In asset management, the potential for more use of these tools should be higher because of the availability of the relevant statistical data on asset performance. This paper presents a more robust risk management process, including use of both formal qualitative risk management planning and quantitative residual risk analysis techniques, which has considerable potential to be used to improve asset performance, decision-making, budget management and timing for major asset management programs.

Key Words: Asset Management, Risk Management, Budget, Maintenance, Forecasting, Stochastic, Correlation

1 INTRODUCTION

The management of a major infrastructure asset such as a power station, smelter or refinery involves understanding and planning for long-term, complex issues as well as market interactions and transactions in which there is always considerable uncertainty in the outcomes. In this paper we define asset management as the 'whole-of-life' operation and maintenance (including major capital refits) of plant and equipment which together form an asset or group of assets.

The traditional approach to asset management is deterministic, that is, operation and maintenance decisions are traditionally based on the forecast of a single particular outcome or operating regime. In reality the level of uncertainty of particular operational events, plant performance, market performance, procurement outcomes etc will vary considerably from forecast results, that is, they are stochastic in nature. Outcomes can be significantly influenced by the prior experience of the asset manager with the plant, the current position in the life cycle of the plant and many other external influences. The major influencers of the level of uncertainty are the risk management or mitigation measures that have been put in place by plant owners, operators or maintainers.

Many effective asset owners consciously and subconsciously adopt a significant level of risk management measures. However, it is not always easy to recognise all the risk management measures that have been put in place. Additionally, it is rare that risk management measures fully remove the risks. For example, risk transfer mechanisms rarely reduce residual risk likelihood or consequence to zero. Asset management teams often have a limited understanding of the residual risk that they carry and conversely, parties such as insurers, financiers, even corporate management teams and/or Boards tend to not fully take into consideration the risk management measures that are in place in determining and making provision for the risk that they carry. This is potentially to the detriment of their success in the markets in which they operate.

Quantitative residual risk analysis is now considerably more accessible as a result of readily available, easy-to-use computer programs and a growing body of knowledge of the use of computer modelling techniques in the asset management context. Quantitative risk analysis has been adopted to a very high level of sophistication in other business realms for a considerable period of time, particularly in the finance and insurance sectors. Useful quantitative analysis must properly account for the qualitative risk management measures that are in place.

Use of a more robust risk management process as part of the asset management process, including use of both formal qualitative risk management planning and quantitative residual risk analysis has considerable potential to improve the outcomes and decision-making in asset management and operational budget preparation. This paper explores this hypothesis and presents some potential tools for use by the asset manager.

2 TYPICAL STRATEGIC PLANNING AND BUDGET PREPARATION PROCESSES

The typical approach to organisational strategic planning and, as part of that process, asset management and budget preparation like all processes that are in need of an overhaul, has evolved over many years and has been driven by the processes and tools employed. The more recent recognition of the priority that the process of asset management needs to be given in the early stages of plant development has begun this overhaul. The next stage of the overhaul of the traditional approach needs to focus on the process of developing or reviewing the annual strategic plans and budgets for the organisation. Typically, this part of the process begins as asset owners and asset managers are approaching the new financial year in response to the requirement to prepare a plan in order to get approvals for activities and financial provisioning for that next year.

Asset managers who are going through this process often have at their disposal an Asset Management Strategy document (hopefully developed during the very early stages of asset life) that is likely to set out technical details of the plant, market position of the company's assets (which allows the plant to quantify production requirements of the plant) and a high-level assessment of market and operating risks. Additionally, an Asset Management Strategy will set out an operating and maintenance regime, which in turn is supported by plant strategies that have been developed and refined using what are regarded today as common maintenance planning tools such as Reliability Centered Maintenance, Root Cause Analysis etc.

The strategic planning and budgeting process is likely to commence with a review of the Asset Management Strategy for changes in external factors such as market conditions or internal factors such as plant condition. This review process is also used to provide a detailed breakdown of activities that need to occur over the period of the strategic plan and budget (often looking little more than one or two years out).

The next stage of the process is the development and costing of a budget that is structured to match activities identified in this latter part of the process. This budget development sometimes utilises a "zero based" approach, which relies on be estimating skills of plant managers and supervisors or, more often than not, is based on last year's spend with some adjustment for the larger activities. It is a very rare that strategic plans or budgets will be adjusted by referring back to the high-level risk assessment in the Asset Management Strategy document or that further risk analysis will be carried out on the budget itself.

A strategic plan and budget thus produced then tends to go through a structured series of reviews and consolidations in preparation for a final approval by senior management or a Board.

The types of less than desirable issues that the approach described above normally leads to are as follows:

- the resultant activities and budget are difficult for senior management to challenge as there is little or no ability to quantify or qualify outcomes if the robustness of the plan is tested and activities are eliminated or budgets reduced. This normally results in adoption of practices where the budget review process becomes a negotiation between asset managers and senior management (normally to the detriment of the asset manager' s position);
- the approved budget becomes both the target for the asset manager to strive for and the financial provision that the organisation makes in its annual funding requirements. This in turn results in poor incentive arrangements for the asset manager, inaccurate reporting of the actual financial position and an inability to effectively fund appropriate operations and maintenance activities following unplanned plant breakdowns, project cost overruns or unforeseen market events;
- the process does not provide any mechanism for optimising asset performance, maintenance spend, risk mitigation strategies or residual risk levels prior to implementation of the plan or during the plan period in response to actual outcomes; and
- the process normally utilises a single market forecast, even though the market forecast itself might have recognised the stochastic nature of market outcomes, as an input to derive a single profitability forecast. This is ineffective use of data since there is no recognition that asset management can, and should, be tailored to match market outcomes.

The utilisation of strategic planning, budget preparation and Asset Management Strategy review processes that incorporate a more thorough qualitative risk analysis and quantitative residual risk modelling have the potential to address each of these issues. Section 3 below describes the terms qualitative risk analysis and quantitative residual risk modelling.

3 RISK MANAGEMENT AND RESIDUAL RISK ANALYSIS

The long duration, unpredictability of plant performance, complex market interaction and exposure of asset management activities to external influences means that at any point in time up until market outcomes have been realised or all costs have been realised, the forecast financial position for an asset will be a range rather than a single number. The uncertainty of the outcome within this range is directly related to the risk profile of an asset and the risk profile in turn is related to the way that asset risks are managed.

In statistical terms, the future financial position of an asset is stochastic in nature, not deterministic. To understand this, it is not difficult for any relatively experienced operations or maintenance personnel to be able to produce an outline probability distribution for the financial outcome of a small maintenance project (say undertaking the overhaul of a major fan or pump) at the end of the project period, by:

- Identifying the extreme best, expected and worst case financial outcomes;
- Dividing the range between best and worst into a number of equal ranges; and
- Estimating the probability of achieving a financial outcome within each of the ranges.

A typical distribution that might be generated in this way is shown in Figure 1.

In Figure 1, point "B" shows the net estimate of $1 million, point "A" the best possible outcome of $930 thousand and point "C" the worst possible case of $1.35 million. There is a 50 per cent probability that the financial position at a particular time will be between 1 million and 1.05 million, a 20 per cent probability that the cost will be between 1.05 and 1.1 million and so on.

The shape of this distribution is a direct reflection of the residual risk and opportunity in a project. The shape will change depending on the risk management measures that are put in place, as well as the procurement methodologies, project delivery methods adopted and other decisions that are reflected directly in the estimate. The same process can be applied to many other asset management activities such as procurement of spares, market impact of maintenance scheduling, market revenues etc.

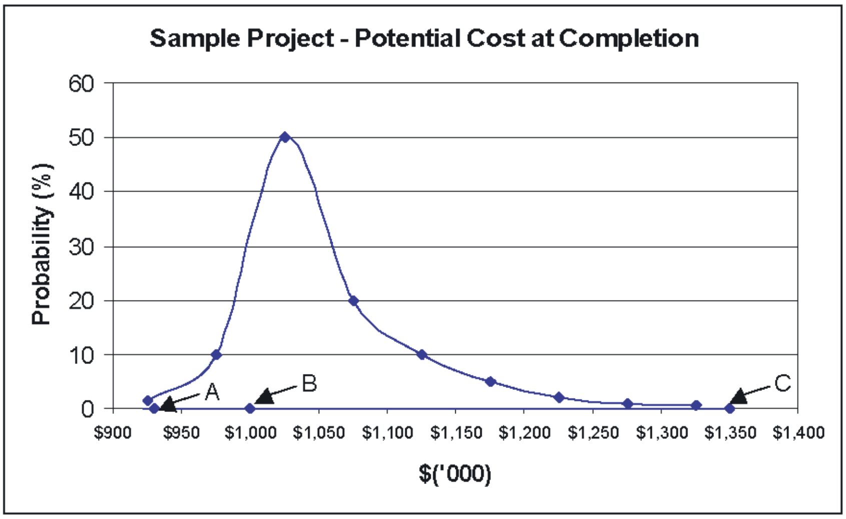


Figure 1 – Distribution of the Forecast Cost at Completion for a Simple Project

3.1 Qualitative Risk Management and Assessment

The risk profile for an asset will depend upon the measures that are put in place to manage risks, including optimising the ability to capitalise on opportunities. Therefore, to measure the potential overall financial outcome for an asset at a particular time, it is necessary to understand:

- all of the potential risks and opportunities;
- how these are being managed;
- whether, after management, there remains some potential likelihood of financial exposure (ie residual risk); and
- if there is residual risk, what potential cost implications it may have.

For example, on a small maintenance project such as the one described above, there will be risks associated with spare parts procurement including both price levels and availability, additional manpower due to difficulties during disassembly or reassembly, or longer term impact due to changes in workplace health and safety requirements each of which can manifest itself in the potential for market losses due to delays. Generally such risks can be mitigated by measures such as a clear delineation of risk under procurement contracts, insurance policies, use of proven maintenance procedures, utilising appropriately skilled and experienced tradespersons and other measures. Notwithstanding these measures, the asset manager can for example be subject to insurance excesses, contract limitations or exclusions or the tradesperson could err in assembly of components. That is, there remains some potential for cost consequences or "residual risk".

The first step in quantifying the cost impact is therefore to assess the risks and risk management measures that exist on such a maintenance project. This is called qualitative risk assessment. There are many ways of undertaking the qualitative assessment including those adopted under international standards and codes of practice.

The basic process involves identifying the risks and opportunities, assessing them generally in terms of likelihood and consequence, identifying the treatment measures that are in place for the risks and opportunities and, where necessary, developing and implementing appropriate, additional risk treatment measures. The process is described in detail for example in AS4360.

Developing comprehensive qualitative risk assessments and optimal management plans is very challenging. Key issues that require to be addressed include:

- establishing a complete picture of all of the risks and opportunities;
- eliminating duplication and overlap;
- recognising the relationships between risks and opportunities, their various sources and root causes; and
- properly recognising all of the treatment measures that are in place including for example generic risk treatment measures (such as appropriately resourced management teams) and measures that offset the effects of risks (such as realising savings elsewhere on a project in response to the manifestation of a risk).

This is a process that most experienced asset managers are familiar with, at least at the high level, as it is the process normally utilised development of an organisation's Risk Register, for the qualification of high level risks in preparing Asset Management Plans or in response to modern Workplace Health & Safety legislative requirements in assessing workplace hazards.

The process is however rarely utilised at lower levels for reviewing projects such as the one used in the example above, for managing the impact of unplanned plant downtime or even for assessing the cost of and prioritising responses to market related event such as abnormally high electricity prices in the electricity market.

This qualitative analysis, which is commonly practiced by asset managers in developing corporate risk registers, is however only the first part of the process we are advocating since it does not entirely eliminate all risks. The second part of the process, and the part that results in the development of several additional useful tools, is the quantitative residual risk analysis.

3.2 Quantitative Residual Risk Analysis and Measurement

The relationship between the qualitative process and the quantitative analysis is illustrated below. As can be seen the outputs of the qualitative process become the inputs to the quantitative process.

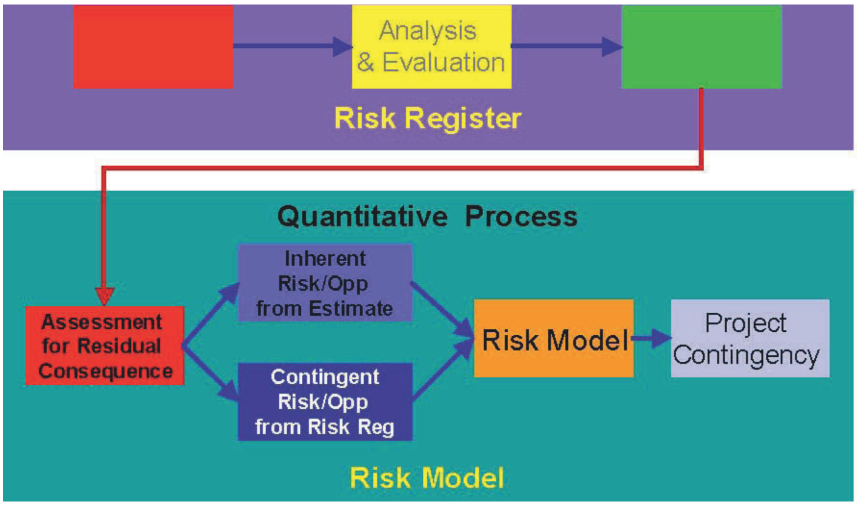


Figure 2 – Relationship Between Qualitative and Quantitative Risk Assessment

Figure 2 shows how the output from the familiar qualitative risk process feeds into the quantitative process which identifies two types of potential sources of residual risk. Inherent risk is the uncertainty in the pricing of the known scope of work, accepting that the asset management strategy is implemented as is contemplated at the time of the assessment; and contingent risks are risk events that may occur during the life of the plan that differ from what has been assumed in the original pricing.

The analysis of the risk profile to develop a quantitative model for potential costs involves using statistical techniques and computer power. The most effective, well recognised of these techniques in this context is Monte Carlo simulation whereby the computer samples very large numbers of potential combinations of risk and opportunity outcomes based on likelihood and consequence parameters provided for each of these (generally in the form of outcome probability distributions for each of these). Subject to proper use, the Monte Carlo modelling and simulation technique is robust, intuitive and widely adopted throughout many spheres of business, particularly in the insurance and finance industries. In developing an asset management strategy and budget (both capital and operating), it would involve:

1. including the potential range of costs for each item of known scope ("inherent risk") usually using a summary of a cost plan or estimate;
2. including the probability of occurrence of each identified risk event and the probable range of costs ("contingent risks"); and
3. simulating potential combinations of the costs of all of these to develop a sufficiently large set of sample total project costs such that, if a completely new set of samples was prepared, the distribution would match the model distribution.

To be accurate, a quantitative risk analysis must obey four rules:

1. It must be complete;
2. The likelihoods and potential cost effects of risks must be valid, meaningful estimates;
3. It must contain no duplication; and
4. The relationships between risks must be reflected in the model, ie independencies and inter-dependencies must be accurately modelled.

The assessment of the potential cost consequence stochastic for each risk may be based on relevant statistics, results on samples of previous annual outcomes or the judgement of asset managers, project teams and/or experts in this type of work.

The likelihood of occurrence of risks must also be assessed and accounted for. Distribution shapes that are commonly used include triangular, uniform, normal, skewed normal, discrete and actual distributions from statistical records.

By examining the potential consequences of each risk, a more detailed view of what might happen on a project overall can be formed.

In undertaking the Monte Carlo simulation, the computer selects a random outcome for each of the risk events and adds these to form a complete scenario. This process is repeated enough times that all possible combinations of events are explored. The density of the total cost outcomes of all of the risk functions then determines the overall forecast cost probability distribution.

Figure 3 below illustrates a forecast budget outcome probability distribution for a sample group of assets.

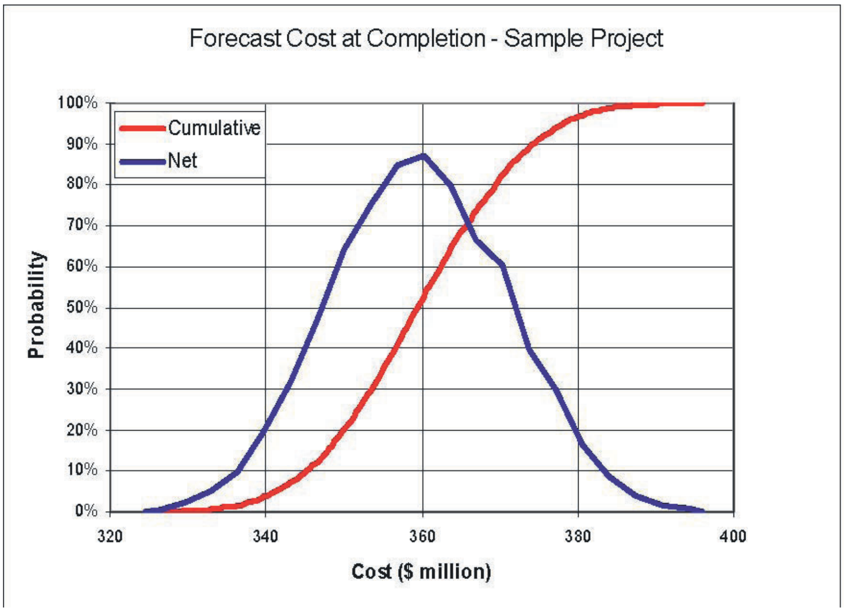


Figure 3 – Cumulative and Net Probability Distributions for a Sample Group of Assets.

This distribution indicates that the forecast budget for the sample assets will be between $327 million and $394 million, that there is a 20 per cent probability that the budget will come in at no more than $350 million (the "P20" figure), a 40 per cent probability that it will come in at no more than $357 million ("P40"), a 60 per cent probability that it will come in at no more than $363 million ("P60") and an 80 per cent probability that it will come in at no more than $370 million ("P80"). It also indicates that the most likely budget outcome is $360 million and that there is a 60 per cent probability that the outcome will be between$350 million and $370 million.

Because the shape of the distribution is a reflection of the risk profile of an asset, the analysis tool can also be used to drive further optimisation of the risk management of the asset. For example, a better managed or scheduled maintenance activity may have a far higher chance of finishing within budget than a less well managed activity. Sensitivity analysis can also be undertaken so that the risks that are having the most potential effect on the project can be identified.

It is a relatively simple conceptual step to recognise that if the same process is utilised for identifying the range and probabilities of market outcomes for the same period that these two analyses can be combined to produce a profitability distribution for the asset. This final step has the potential to produce a most invaluable tool that allows testing and therefore optimisation of operations and maintenance activities to match the variability of anticipated market outcomes.

4 USES FOR RISK ANALYSIS OUTCOMES IN DECISION-MAKING

Assuming that the qualitative risk management work was thoroughly undertaken and the residual risk was accurately assessed, the recognition of the stochastic nature of the potential financial outcomes and an understanding of the various parameters that can affect those outcomes can mean that these models can be useful tools for optimising decision-making and in guiding asset management activities.

4.1 Difficulties in accurate modelling for decision making

The assumption that the risk models are accurate are sweeping and can be found wanting. The key issues in preparing accurate risk models include:

1. Failure to account for the ability of teams to overcome risks in ways not fully understood when the model is built. For example, a risk cost may be offset by cutting a cost or undertaking different work on an entirely un-related component of the project;
2. Incomplete models;
3. Failure to recognise correlation, between components of projects and/or between different projects within a portfolio;
4. The uniqueness of risks and the lack of available data; and/or
5. The over-provision for contingencies at corporate level as a result of a failure to recognise the combined "portfolio" effect of individual project stochastics (discussed below).

Ultimately, risk analysis involves both endeavouring to comprehend what may happen in the future and simplifying what may happen to the point where a model can be prepared in a sensible amount of time. When considering asset management activities, this simplification process is assisted where there are accurate and complete plant history records available which provide a guide to the range of outcomes and risks and data that can assist in building appropriate risk distributions. Such historical data is commonly available for today's asset managers utilising computer maintenance management systems.

Regardless of some of the difficulties faced in preparing accurate risk models, the use of the residual risk stochastic will be of assistance in providing quantitative support for adopting reasonable contingencies, deciding whether to proceed with or modify certain asset maintenance activities or in recognising the potential for maintenance costs to overrun notwithstanding contingencies and so on.

4.2 Portfolio Effects

In making decisions about an individual asset, many business entities will be considering other components of their business, particularly other assets that they have on the books. Portfolio theory in risk analysis is another area that has been very widely explored in the finance industry.

The key outcome of an analysis of portfolios is that the combined levels of risk of all of the assets will be less than the arithmetic sums of the component assets, assuming that there is not perfect correlation between the various assets. For example, the probability of 5 uncorrelated assets all achieving budget outcomes at a P65 cost or greater is $(0.35)^5 = 0.005$, or less than half a per cent.

The corollary of this is that a portfolio manager can have an overall "budget provision" that is smaller than the arithmetic sum of the budget provisions required for the individual assets, whilst effectively having sufficient budget provision for each asset.

An example of this effect in practice would be some of the local Electricity Generation owners who have multiple power station sites operating in different parts of the market requiring quite different operating regimes to be implemented at each. These assets can have quite a low correlation and therefore this portfolio effect can be very useful for allocating appropriate budget provisioning.

5 ASSET MANAGEMENT AND RISK ANALYSIS OUTCOMES

The above discussion focuses on the uses of forecast stochastics in decision-making.

The stochastics also serve to illustrate in an obvious way the differences between targets, forecasts and financial provisions. During the development of an asset budget and strategic plan, it is typically a challenge for management at all levels to filter cost information that is available to understand the potential budget outcomes, accounting for such factors as the optimism of the asset management team, the risks that may yet arise and the potential for costs and/or incomes to change before the plan is complete.

Until the plan is complete and all financial accounts have been settled, there will be some cost and/or price uncertainty. Asset managers are duty-bound to strive for targets (both cost and price, depending on the commercial arrangements). Their regular reports to senior management need to provide accurate information on forecasts – not necessarily the same as the targets. Senior management also needs to be able to ensure that sufficient funds are available in the event of overruns. Figure 4 below illustrates how the cost figures might look at some point through the plan life for the group of assets used in section 3.2 above.

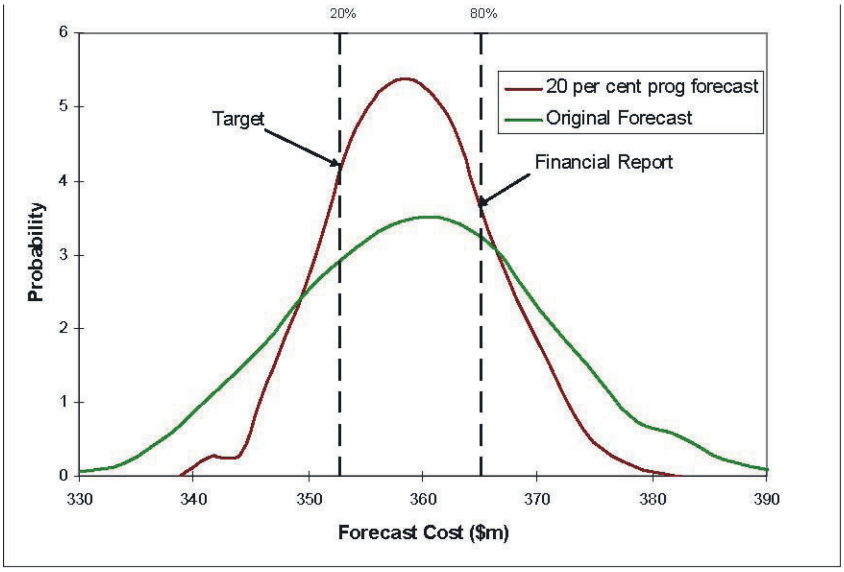


Figure 4 – Forecast Outcome Figures for 35% Complete Budget Plan

In this example, it may be appropriate for senior management to set the target cost at the P20 figure of $353 million. It stands to reason that the forecast expected cost is the "most likely" cost of $358 million. (In most computer-generated Monte Carlo simulation modelling, the use of the most likely figure or mode can be problematic because of the way the models are presented. A more reliable measure should typically be around the median or mean, which are typically close together in asset cost forecast models). The senior manager could make financial budget provisions at the P80 figure of $365 million, although this might need adjustment at portfolio level as discussed in section 4.2 above.

For appropriate targets to be reflected in forecast distributions, it is essential that all opportunities are properly reflected in the qualitative assessment and then as inputs into the quantitative modelling. It is also essential that due account is taken of risk treatment measures including less obvious responses to risks that may arise such as initiatives taken by good asset management teams that are not formally recorded in the qualitative assessment process. Many quantitative models inadequately reflect "upside" cost outcomes as the modelling team tend to concentrate more on the potential downside. Setting stretch targets requires that the better outcomes are given equal attention to the downside outcomes.

A related issue with the overhaul of major assets is the uncertainty in overhaul project durations. For owners and contractors, a considerable proportion of the financial risk will be in the potential for delays resulting in time-related costs. For example, productivity losses due to plant/asset access restrictions. Clearly, the investigation of the uncertainty on the durations of overhaul projects, individual tasks and the relationships between tasks will assist in providing more meaningful, accurate cost forecasts.

As asset management activities progress through the period of the Strategic Plan, the differences between the targets, forecasts and financial provision figures will narrow.

5.1 Budget Issues for Senior Management

Section 2 of this paper addresses a number of issues the Senior Management face during the review of asset strategy plans, corporate strategic plans and associated budgets. The discussion below describes how the authors see the adoption of a quantitative risk modelling approach will assist senior management to address these issues.

The first issue identified was that activities and budget could be difficult for senior management to challenge as there was little or no ability to quantify or qualify outcomes if/when the robustness of the plan is tested and activities are eliminated or budgets reduced. If the asset plan and associated budgets are developed utilising the risk analysis process set out above, the asset manager is provided with an ability to analyse the residual risk if a particular activity is modified or eliminated altogether, and to quantify any argument for the inclusion of particular activities. In other words the outcomes of this process allows the asset manager to demonstrate from a cost or residual risk perspective why particular maintenance levels or projects are being undertaken and allows senior management to understand the risks they must accept if such activities are not undertaken. It also allows both parties to identify and prioritise activities and tasks that should be undertaken if, at any time through the plan period, budget outcomes are forecast to be better than expected.

The second issue was the issue of the traditional approach to developing a budget resulting in budget targets matching financial provisions that are made by the organisation. Section 5.1 above clearly demonstrates the flexibility that the proposed approach yields in relation to both setting stretch targets for asset management and making adequate budget provision to cater for potential risks of cost overrun at the same time.

Adoption of the process obviously leads to the development of a set of tools that allow a significant level of "what if" analysis to facilitate the optimisation of asset performance and maintenance spend.

With the ability to combine the stochastic outputs from both the asset management strategy and the market outcome analysis, the proposed process provides an unparalleled tool for linking asset management decision-making with market priorities. This particular area is one that has traditionally relied upon the expertise and experience of the asset manager who has never been able to rely upon quantifiable analysis to provide assistance in this kind of decision-making.

6 CONCLUSION

Management of major assets involve complex, long life transactions that generally involve considerable degrees of uncertainty. The uncertainty is a function of the risks on the asset and the measures in place to manage that risk. All financial stakeholders will be affected by the uncertainty and therefore these risk management measures. Measures such as risk mitigation or transfer are rarely 100 per cent effective; there is generally some likelihood that the measures will prove ineffective and that some cost can be incurred.

Qualitative risk analysis of high level risks has become a common practice amongst asset owners particularly in relation to development of a corporate risk register. Quantitative risk models, generated by easy-to-use, readily available desktop computer applications are now used to greater or lesser degrees by entities to quantify some of the market risks they face but the same techniques are rarely utilised by asset managers to quantify the risks they face in delivering a strategic plan and budget. Provided these quantitative risk models can be developed to reflect the risk mitigation measures that are in place they can be used to inform decision-making in costing of works and understanding potential financial exposures. They can be used in asset strategies to set targets, separate targets from forecasts and understand the potential for costs to overrun forecasts.

The models can also be used at portfolio level (programme, site or corporate level) to inform decision-making at that level. Portfolio effects, long understood and beneficially used in the finance industry, can also be understood and beneficially used in management and decision-making.

It is difficult to develop accurate models but useful models have been developed and successfully used by Evans & Peck in the management of a number of major infrastructure projects and programs. It is essential to recognise the potential for weaknesses in models and, in making use of them, to subject them to adequate scrutiny and to limit reliance on them consistent with informed judgement of their accuracy. Nevertheless, quantitative models of forecast outcomes, with all their weaknesses, are in widespread use in finance and other areas of industry, and have considerable potential be used with equal benefit in the asset management arena.

7 REFERENCES

[1] Trueman P (2004) Capital Works Decision Making Using Risk Management Techniques Proceedings of the Australian Risk Engineering Society Conference, November 2004.

[2] Bernstein PL (1996) Against the Gods: The Remarkable Story of Risk. United States of America, John Wiley & Sons, Inc.

LOAD-INDEPENDENT CONDITION ASSESSMENT OF GEARS USING KOLMOGOROV-SMIRNOV GOODNESS-OF-FIT TEST AND AUTOREGRESSIVE MODELING

Zhan YM and Mechefske CK

Department of Mechanical and Materials Engineering, Queen's University, Kingston, Ontario, Canada K7L 3N6

E-mails: zhan@me.queensu.ca; chrism@me.queensu.ca

Abstract: A novel technique for detection of gearbox deterioration is proposed in this study. The proposed technique makes use of a time-varying autoregressive model and establishes a compromised autoregressive model based on healthy-state gear motion residual signals under varying load conditions and employs the Kolmogorov-Smirnov goodness-of-fit test statistic as a measure of gear condition. The order of the time-varying autoregressive model is selected by using a proposed fully automatic and highly effective model order selection technique with the aid of hypothesis tests. Validation of the proposed technique is carried out by using both simulated and real entire lifetime gear vibration signals and demonstrates that the proposed technique possesses appealing effectiveness in identifying the optimum autoregressive model order for robust gear condition detection under varying load conditions.

Key Words: Vibration Monitoring, Gearbox, Condition Indicator, Autoregressive Model, Kolmogorov-Smirnov Goodness-of-Fit Test, Hypothesis Tests.

1 INTRODUCTION

A significant surge of research achievements on the condition monitoring of rotating machinery has been reported in the latest two decades. Particular attention should be given to the pioneering work by McFadden [1, 2] who proposed a highly efficient vibration monitoring technique for gearing systems using Hilbert transform based phase modulation (PM) and amplitude modulation (AM). However, this technique involves a strong subjective determination of the dominant meshing frequency and accompanying sidebands. Inappropriate selection of dominant meshing frequency may lead to poor quality detection of the target gear. Such a disadvantage was denoted in some recent research [3, 4].

A recent study by Wang [4] proposed a narrowband structural resonance demodulation based technique which to some extent is able to provide an earlier alert of incipient tooth crack in comparison with the aforementioned PM method by McFadden. Nevertheless, this technique is in need of cautious user provided identification of the frequency band which contains the dominant structural resonance. Such a necessity considerably fetters the real-time application of this technique.

It should also be noted that time-frequency analysis, in particular the wavelet transform, has been an intense research focus in the fault detection and diagnosis of rotating machinery in the latest decade [5−7]. However, the expert based explanation of the resulting time-frequency representations limit the application of these techniques.

There remains a need for an alternative technique in the time-domain which is able to capture the non-stationarity of a signal and show insight into its dynamics, but simultaneously be time-efficient. Kalman filter based time-series analysis is such a candidate technique. It possesses desirable properties but has not yet attracted sufficient attention in the fault detection and diagnosis of rotating machinery. Furthermore, it is worthwhile to note that in recent years the Kolmogorov-Smirnov (K-S) goodness-of-fit (GOF) test has been found to be an extremely powerful tool in the condition monitoring of rotating machinery. Some preceding research using the K-S test has been carried out in the fault detection and diagnosis of gears and bearings and other rotating components [8−10].

Since a desirable detection model must be able to best represent the healthy vibration signals under varying load conditions and be most sensitive to any abnormality caused by a defective condition, the Kalman filter based autoregressive model (Appendix A) is employed in this study to capture the dynamics of healthy vibration signals under varying load conditions, while the K-S GOF test statistic (Appendix B) is used to provide a quantitative indication of the change in the condition of the monitored object. A novel compromised autoregressive (AR) modeling technique with the aid of hypothesis tests will be proposed in this study, which is followed by a simulation analysis using artificially generated full lifetime gear vibration signals and an experimental analysis using real gearbox entire lifetime vibration signals.

The paper is organized as follows. Section 2 presents the theoretical elements of the proposed gear condition detection technique. Section 3 presents a simulation analysis using simulated entire lifetime gear vibration signals. Section 4 presents an experimental analysis using real entire lifetime gear vibration signals. Discussions and conclusions are presented in Section 5.

2 THEORETICAL DEVELOPMENT

2.1 Description of Model

This section proposes a fully automatic condition detection technique for gearbox condition monitoring using a Kalman filter based AR model with the aid of hypothesis tests for the selection of optimum model order to achieve the predominant objective, i.e. consistent normality of the AR model residuals for healthy gear motion residual (GMR) signals under varying load conditions.

In the practical operation of the proposed technique, the GMR signals are divided into two groups: the control group and the treatment group. The control group GMR signals that are collected with the target gear in a healthy condition and under varying load conditions are exclusively used for establishing a compromised AR model fit. It needs to be emphasized that only the healthy condition GMR signals collected under the two extreme load conditions (lower and upper load limits) are needed for establishing a compromised AR model fit. The treatment group GMR signals contain the entire lifetime GMR signals of a gearbox running from new to failure and are used for validation of the established compromised AR modeling technique.

Assume that there are N_1 GMR signals collected under the lower limit load condition and N_2 GMR signals collected under the upper limit load condition available for use. Thus, the control group of GMR signals is of size N_1+N_2. Each GMR signal contains n sample points equally spaced over a complete revolution (0° to 360°) of the target gear. Under a certain AR model order p within the considered order range $[p_s, p_e]$, the adaptive fashion of the Kalman filter yields a set of AR model coefficients a^l_{ijk}, i=1…N_1, j=1…n, k=1…p for the lower limit load condition where the superscript l stands for lower load condition, and another set of AR model coefficients a^u_{ijk}, i=1…N_2, j=1…n, k=1…p for the upper limit load condition where the superscript u stands for upper load condition, where $a^l_{i\bullet\bullet}$ or $a^u_{i\bullet\bullet}$ stands for the AR model coefficient matrix of dimension $p\times n$ over the complete revolution of each GMR signal among N_1 or N_2, and $a^l_{ij\bullet}$ or $a^u_{ij\bullet}$ stands for the AR model coefficient vector of dimension $p\times 1$ at each time instant in each GMR signal among N_1 or N_2.

In the case where phase information is not available, an averaging operation

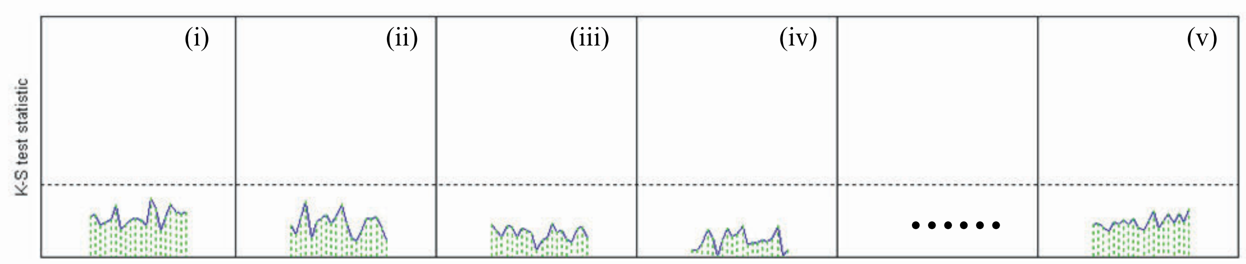
$$\bar{\bar{a}}_{\bullet} = \frac{\sum_{i=1}^{N_1}\sum_{j=1}^{n} a^l_{ij\bullet} + \sum_{i=1}^{N_2}\sum_{j=1}^{n} a^u_{ij\bullet}}{(N_1 + N_2)\times n} \tag{1}$$

is carried out, where $\bar{\bar{a}}_{\bullet} = \bar{\bar{a}}_k$, k=1…p. Note that the proposed $\bar{\bar{a}}_k$ differs from the AR model coefficient vector a_k, k=1…p described by Wang and Wong [11] which is estimated on the assumption of stationarity by the conventional Yule-Walker method. Eq. (1) will be used throughout this study. The obvious reason for using this constant state vector is that the healthy condition GMR signals that were used to construct the compromised AR model may not be in phase with future GMR signals.

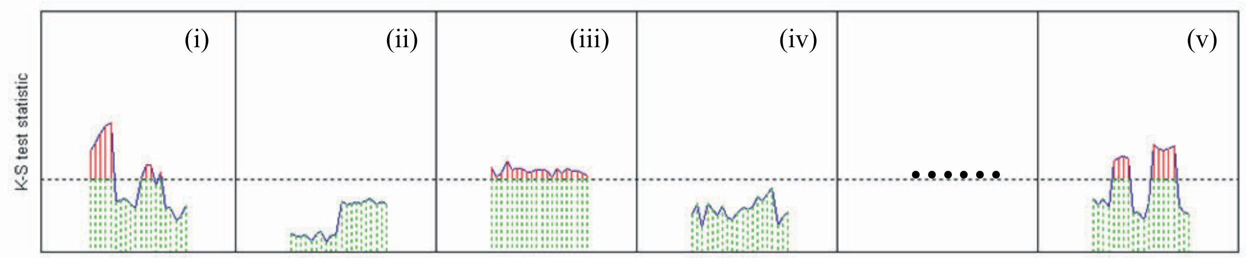
The second signal processing operation is based upon an important previous finding that the AR model order is the decisive control factor which determines whether or not a compromised AR model that is robust with respect to varying load conditions can be established [12]. The K-S test statistic is applied to the AR model residuals of each GMR signal, as a quantitative gear condition indicator. In practical implementations, two types of AR model order within the considered range $[p_s, p_e]$ may be present and are defined as below:

(a1) Type I orders within $[p_s, p_e]$ that enable the compromised AR model fitting procedure to generate normally distributed AR model residuals for all GMR signals in the control group;

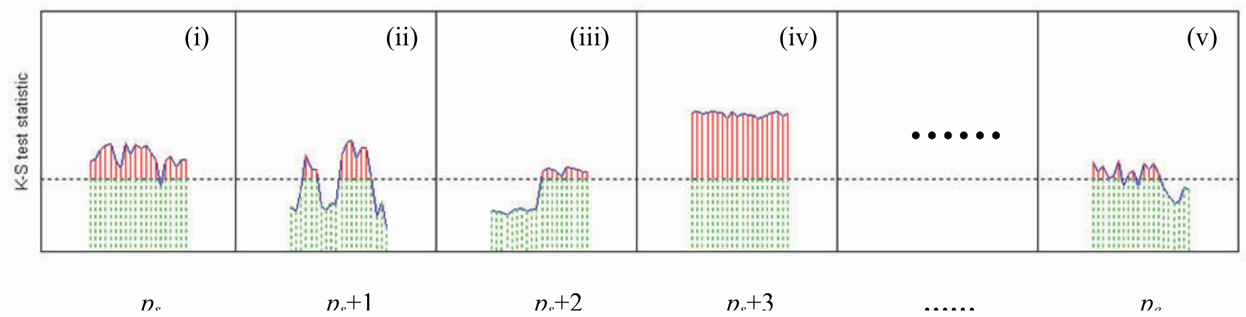
(a2) Type II orders within $[p_s, p_e]$ that enable the compromised AR model fitting procedure to generate normally distributed AR model residuals only for part of the control group GMR signals.



a) The first situation (b1) which all the orders in $[p_s, p_e]$ are Type I



b) The second situation (b2) which the orders in $[p_s, p_e]$ are either Type I or Type II



c) The third situation (b3) which all the orders in $[p_s, p_e]$ are Type II

Figure 1. The three situations that contain the Type I and Type II AR model orders within the considered order range $[p_s, p_e]$. The horizontal dotted line in each subplot denotes the K-S critical value.

Consequently, three possible situations may take place:

(b1) All orders within $[p_s, p_e]$ are Type I, as shown in Figure 1(a);

(b2) Some orders within $[p_s, p_e]$ are Type I, while others are Type II, as shown in Figure 1(b);

(b3) All orders within $[p_s, p_e]$ are Type II, as shown in Figure 1(c).

In terms of our experimental investigation, the first situation (b1) is rare, while the second and the third situations, (b2) and (b3), are often encountered. It is natural to think that priority should be given to the Type I orders when the first or second situation, (b1) or (b2), appears. Under such circumstances, the optimum AR model order should be exclusively chosen from the Type I orders, while the remaining Type II orders will not be considered if the second situation (b2) is encountered. The proposed procedure of selecting the optimum AR model order from Type I orders is summarized below:

(c1) Search for the order p_{lm} which gives the least mean of the K-S test statistics over all control group GMR signals;

(c2) Calculate the variance V_{lm} of the K-S test statistics over all control group GMR signals under the order p_{lm};

(c3) Search for the order p_{lv} which gives the least variance V_{lv} of the K-S test statistics over all control group GMR signals;

(c4) Compare V_{lm} with V_{lv} using the Bartlett's test under the null hypothesis H_0 and the alternative hypothesis H_a:

$$H_0: V_{lm}=V_{lv}$$

$$H_a: V_{lm} \neq V_{lv}$$

at a significance level of 5%;

(c5) If the null hypothesis H_0 is accepted, the optimum AR model order $p_{optimum}$ takes the value of p_{lm}. Otherwise, $p_{optimum}$ takes the value of p_{lv}.

By following this procedure, the order p_{lm} with the least mean of the K-S test statistics is selected if its variance is not statistically different from that of the order p_{lv} with the least variance of the K-S test statistics. Otherwise, the order p_{lv} with the least variance is selected. Therefore, the normality of AR model residuals of the control group GMR signals collected under the two extreme load conditions is always guaranteed by the determined $p_{optimum}$, while in the mean time it possesses the least variance.

On the other hand, when the third situation (b3) occurs, the objective of searching for $p_{optimum}$ is to find an order from Type II orders which gives the least violation against the normality assumption of the AR model residuals of each control group GMR signal in the sense that, over the full control group GMR signals, the K-S test statistics beyond the K-S critical value remain at the least mean level. Therefore, our interest is in the analysis of the K-S test statistics which exceed the critical value under a certain Type II order and thus are termed K-S outliers, whereas other K-S test statistics that are below the critical value, under the present order, are not considered. As such, a different procedure of searching for $p_{optimum}$ from Type II orders for the third situation (b3) is proposed below:

(d1) Search for the order p_{lm} which gives the least mean M_{least} of the K-S outliers;

(d2) Search for all the orders whose means of the K-S outliers are statistically equal to M_{least} using the Satterthwaite's t'-test under the null hypothesis H_0 and the alternative hypothesis H_a:

$$H_0: M_{least} = M_i$$

$$H_a: M_{least} \neq M_i$$

at a significance level of 5%, where M_i is the mean of the K-S outliers of an order p_i in $[p_s, p_e]$;

(d3) Let a set $P=\{p_j, j=1\ldots k\}$ store all the orders found in (d2);

(d4) Calculate the area A_j between the curve of the K-S outliers and the horizontal line of the K-S critical value for each p_j in set P, as shown in the area covered by vertical solid lines in each subplot of Figure 1(c);

(d5) Compare all the A_j's and search for the order p_{la} in the set P which has the least area A_j;

(d6) The optimum AR model order $p_{optimum}$ takes the value of p_{la}.

In the above procedure, the order with the least mean level of K-S outliers and the least area beyond the K-S critical value can be determined, which is optimum in the sense that the violation against the normality assumption of AR model residuals remains at the lowest level. The application of the Satterthwaite's t'-test, in contrast to the conventional t-test, is used in view of a realistic consideration that the homoscedasticity of the K-S outliers under different Type II orders cannot be guaranteed. This is known as the Behrens-Fisher problem, whereas the homoscedasticity is a major assumption for the conventional t-test.

2.2 Some notes

Note that in the second processing, for the situations (b1) and (b2), comparing to other orders within $[p_s, p_e]$ the least variance of the K-S statistics of the selected $p_{optimum}$ is ensured, which implies that the mean M_{N_1} of the K-S statistics of the N_1 control group GMR signals corresponding to the lower limit load condition cannot differ much from the mean M_{N_2} of the N_2 control group GMR signals corresponding to the upper limit load condition. Otherwise, the variance will be increased. Hence, the load-independence can be achieved to the best available extent.

However, the case is somewhat different when the third situation (b3) comes forth since all orders within $[p_s, p_e]$ are Type II. Thus, the primary objective is to search for an order to bring the least violation against the normality assumption of the AR model residuals rather than the equality between the M_{N_1} and M_{N_2}. This is done to avoid the realistic possibility that, for the third situation (b3), the equality between the M_{N_1} and M_{N_2} which takes priority over the normality assumption may result in the selection of an order which strongly violates the most desired normality assumption

Further, when the normality assumption is ensured to the best available extent in the sense that some K-S test statistics of the control group GMR signals only slightly exceed the K-S critical value, the varying load induced difference between the M_{N_1} and M_{N_2} cannot be large in comparison with the fault-induced variations of the K-S test statistic in its future application because the K-S critical value is a small value. Under the same principle, a similar possible case may happen to the second situation (b2), as shown in Figures 1(b-ii) and 1(b-iii), where the order p_s+1 is thus preferred instead of p_s+2.

To ensure the normality of AR model residuals as best as we can during the establishment of the compromised AR model fit using healthy condition GMR signals will reveal the presence of incipient defective gear conditions. This is due to the fact that a gear defect usually gives a slight deviation from normality in the AR model residuals at an early stage of development, although it is sometimes at the expense of deteriorating the equality requirement for the M_{N_1} and M_{N_2}.

2.3 Definition of gear condition indicators

Since our interest in the condition monitoring of gears using the condition detection technique proposed in Section 2.2 is to detect the deviation of AR model residuals from the normality assumption, we modify the conventional K-S test statistic to be

$$MKS = (KS - KS_{cv})I(KS > KS_{cv}) \tag{2}$$

where KS stands for the original K-S test statistic at a significance level of 5%, KS_{cv} stands for the K-S critical value, MKS stands for the modified K-S test statistic, and $I(\bullet)$ is a sign function. By means of the Eq. (3), the deviation from the normality assumption of AR model residuals determined from healthy gear conditions can be detected.

For the kurtosis measure, the normality is achieved only when it is equal to 3, beyond which or below which the distribution is termed leptokurtic or platykurtic, respectively. Thus, the modified kurtosis measure to detect the deviation from normality is

$$MKT = |(KT - 3)| I(KT \neq 3) \tag{3}$$

where KT stands for the original kurtosis value, MKT stands for the modified kurtosis value, and $|\bullet|$ stands for the absolute operation.

3 SIMULATION ANALYSIS

3.1 Signal model

This study proposes more realistic baseline functions of simulated gear vibration signals under healthy and defective conditions, which are given by

$$x(t) = \sum_{m=0}^{M} (1 + c_i^k) X_m^k [1 + a_m(t)] \cos\{2\pi m T f_s t + \phi_m + b_m(t)\} \tag{4}$$

and

$$x(t) = \sum_{m=0}^{M} (1 + c_i^k) X_m^k [1 + \tilde{a}_m(t)] \cos\{2\pi m T f_s t + \phi_m + \tilde{b}_m(t)\} + h_3 z(t) \tag{5}$$

where c_i^k , i=1,…,L is a random coefficient series corresponding to a total of L vibration signals collected at the k-th load condition, $(1 + c_i^k) X_m^k$ is the modified amplitude of the m-th meshing frequency at the k-th load condition. When a localized gear fault occurs, localized impacts may modify the healthy condition modulation functions $a_m(t)$ and $b_m(t)$ expressed by

$$a_m(t) = \sum_{n=1}^{N} A_{mn} \cos(2\pi n f_s t + \alpha_{mn}) \tag{6}$$

and

$$b_m(t) = \sum_{n=1}^{N} B_{mn} \cos(2\pi n f_s t + \beta_{mn}) \tag{7}$$

to be $\tilde{a}_m(t)$ and $\tilde{b}_m(t)$ expressed by

$$\tilde{a}_m(t) = a_m(t) + h_1 d(t) \tag{8}$$

and

$$\tilde{b}_m(t) = b_m(t) + h_2 d(t) \tag{9}$$

where h_1 and h_2 are coefficients which determine the magnitude of the localized impact induced additional amplitude modulation and phase modulation. At the same time, localized impacts may also excite structural resonances. Using the same notation as that used by Wang and Wong [11], we denote the impact-induced resonant vibration in one revolution of the monitored gear by $z(t)$ which is expressed by:

$$z(t) = d(t)\cos(2\pi f_r t + \eta_r) \tag{10}$$

where $d(t)$ is the envelope (modulating) function of the resonant vibration, f_r the resonance frequency (carrier frequency), and η_r the corresponding initial phase. The envelope function $d(t)$ takes the normalized form of a modified log-normal distribution function, which is given by [4]

$$d(t) = d_0(t) / \max |d_0(t)| \tag{11}$$

where $d_0(t)$ is given by

$$d_0(t) = \frac{1}{t\sigma_{impact}\sqrt{2\pi}} e^{-\frac{(\ln t - \mu_{impact})^2}{\sigma_{impact}^2}} \tag{12}$$

where σ_{impact} and μ_{impact} stand for modified log-normal distribution variance and mean which determine the shape of the envelop function of the resonant vibration.

Table 1 shows the basic information of this simulated test. As shown in Table 1, the gear of interest has 27 teeth, a maximum number of considered gear meshing frequencies of 6, a shaft speed of 40 Hz, and a sampling frequency of 40 kHz. Each TSA signal has a length of 1000 sample points and thus each extracted GMR signal.

Table 1

General description of the simulated test

T	27
M	6
f_s (Hz)	40
S	6 (one healthy and five defective)
K	7
L	5
μ_{impact}	-8.5172
σ_{impact}	1
f_r (shaft order)	121
η_r	$-\pi/2$
N	3

Sampling frequency (kHz)	40
Length of each TSA/GMR signal	1000 data points

3.2 Condition detection

This section also proposes some alternative condition indicators and carries out a comprehensive comparison. For convenience, abbreviations MKS_1, MKT_1, MKS_2, and MKT_2 are used, where MKS_1 is the proposed technique in this study, MKT_1 is a kurtosis based alternative indicator to MKS_1, MKS_2 is the modified K-S statistic applied to gear motion residual signal, and MKT_2 is the conventional gear motion residual signal kurtosis based method.

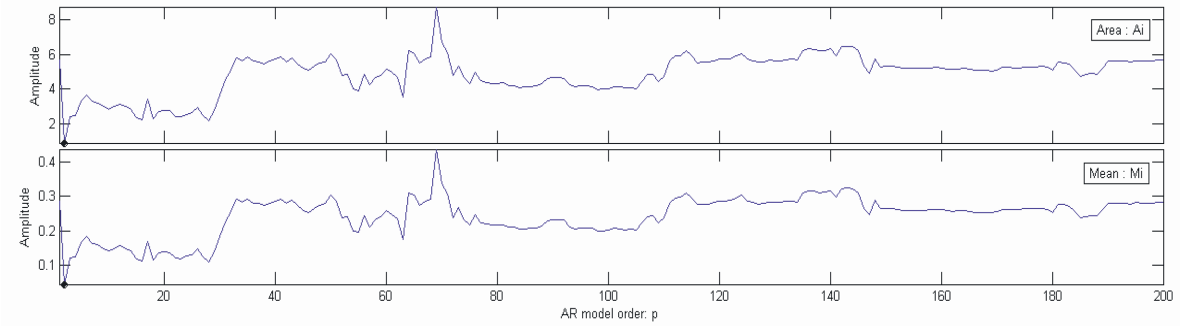
Figure 2 shows the AR model order selection for MKS_1 and MKT_1 using the control group GMR signals, which has 40 GMR signals in total. The optimum orders are selected from Type II orders for both MKS_1 and MKT_1. The selected optimum orders are 2 for MKS_1 and 182 for MKT_1. Note that in each of Figures 2(a) and 2(b), the area curve A_i and the mean curve M_i have identical trace. This is because, at each order, the number of K-S outliers is constant over the AR model order horizon.

As shown in Figure 2(b), the Satterthwaite's t'-test at a significance level of 5% found a few orders for MKT_1 which have statistically equal means of K-S outliers to that of the $p_{optimum}$.

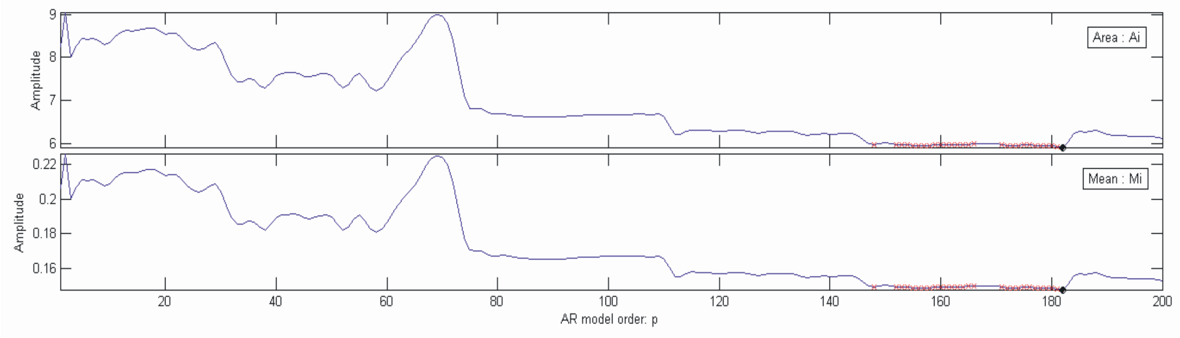
Figures 3(a) and 3(b) show the state detection using the optimum AR model orders of MKS_1 and MKT_1 found in Figure 2. The four gear condition indicators all give the first alarm at the GMR signal #36 when the first defective condition GMR signal is encountered.

Within the healthy condition of the gear, i.e. GMR signal interval [1:35], the proposed MKS_1 presents the least violation against the normality assumption as denoted by the downward arrow mark in Figure 3(a), while the MKS_2 ranks second as denoted by the downward arrow mark in Figure 3(c).

In addition, by comparing the trace of MKS_1, as shown in Figure 3(a), with that of MKT_1, as shown in Figure 3(b), it is found that the kurtosis-based gear condition indicator MKS_1 demonstrate strong varying-load induced fluctuations in contrast with the K-S test statistic-based condition indicator MKT_1. For instance, within the GMR signal interval [71:105], the variability of MKS_1 caused by varying load conditions as denoted by the upward arrow mark in Figure 3(a) is much less than that of MKT_1 as denoted by the upward arrow mark in Figure 3(b). The same comparison and conclusion can be applied to and

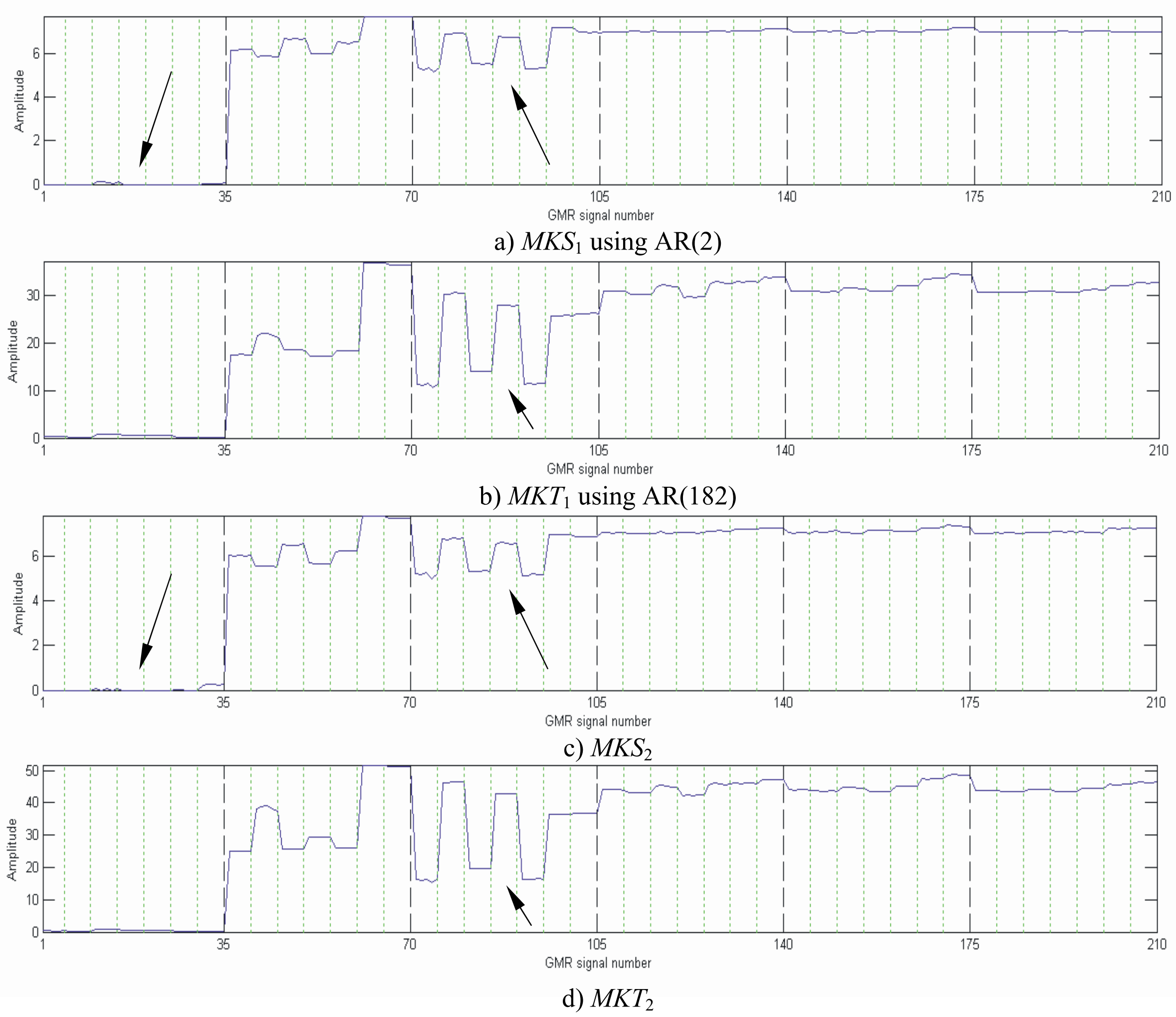


a) $p_{optimum}$=2 for MKS_1, selected from Type II orders



b) $p_{optimum}$=182 for MKT_1, selected from Type II orders, where the orders marked by cross markers are the orders with means of K-S outliers which are statistically equal to that of the $p_{optimum}$ via the Satterthwaite's t'-test at a significance level of 5%

Figure 2. AR model order selection, where the order marked by a dot in each subplot is the selected order $p_{optimum}$.



a) MKS_1 using AR(2)

b) MKT_1 using AR(182)

c) MKS_2

d) MKT_2

Figure 3. Gear condition detection using simulated clean lifetime GMR signals. The dotted lines denote the boundary of each load condition. The dashed lines denote the boundary of each gear condition.

drawn from MKS_2 and MKT_2 as denoted by the upward arrow marks in Figures 3(c) and 3(d).

In general the proposed MKS_1 outperforms other gear condition indicators due to its most significant first alarm for incipient fault, least violation against the normality assumption of AR model residuals within the healthy period of the investigated gear, and the least varying-load induced fluctuations.

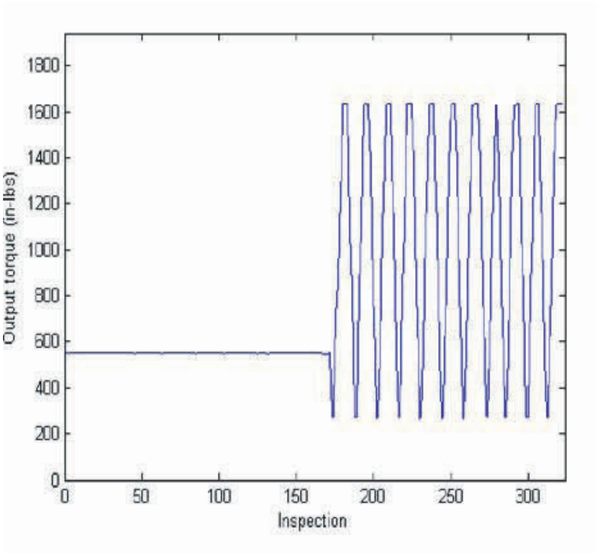
3.3 Optimality checks

To validate the optimality of the $p_{optimum}$ selected by the proposed compromised AR model fitting technique, gear condition detection using MKS_1 and MKT_1 under alternative orders with similar means, M_i's, are carried out. (Note: Not necessarily statistically equal means because, for instance, for MKS_1 as shown in Figure 2(a), the selected order p=2 does not have statistically equal means from other orders).

Table 2 shows a numerical analysis of the mean of K-S outliers expressed by

$$\sum_{i=1}^{35} MKS_{i} I(MKS_{i} > 0) \Big/ \sum_{i=1}^{35} I(MKS_{i} > 0) \tag{13}$$

for the performance of MKS_1 within the healthy GMR signal interval [1:35] under alternative orders 1, 3, 16, 18, 23, and 28, respectively. The minimum is achieved at the selected order p=2, which indicates that the normality assumption of AR model residuals within the healthy gear condition is best achieved by the selected $p_{optimum}$. Therefore, it confirms the optimality of the $p_{optimum}$=2 for the proposed MKS_1.



a) The output torque level switches from a constant 555 in-lbs at the 172nd inspection to a sinusoidal load condition at the 173rd inspection

b) Post drive gear with seven broken teeth

Figure 4. Load condition and the post drive gear of TR#14

A similar numerical analysis was carried out for MKT_1 under alternative Type II orders, which, for simplicity, was not presented here. The result indicates that the selected optimum order p=182 results in the minimum mean of K-S outliers within the healthy period of the gear of interest. Therefore, it confirms again the capability of the proposed compromised AR model fitting technique in the selection of a robust model order under varying load conditions.

Table 2

Comparison between the $p_{optimum}$ and alternative orders in gear condition detection using MKS_1

Order p	Mean of K-S outliers within the healthy gear condition [1:35]
2 ($p_{optimum}$)	6.549E-02
1	1.650E-01
3	7.430E-02
16	9.092E-02
18	8.684E-02
23	8.752E-02
28	7.492E-02

4 EXPERIMENTAL ANALYSIS

4.1 General description of TR#14

Full lifetime data of test-run #14, provided by the Applied Research Laboratory (ARL) at the Pennsylvania State University, will be used in this study. Researchers used various combinations of motors, equipment, and sensors on its

Mechanical Diagnostic Test Bed (MDTB) and ran the MDTB to failure, collecting the resulting data and carefully documenting their observations [13, 14].

There were a total of 323 data files collected during the lifetime of TR#14. The gearbox in TR#14 had a 70:21 tooth ratio and was driven at the 100% nominal output torque level of 555 in-lbs for 95 hours in which 172 data files were collected. Subsequently, during the sinusoidal load condition, which varied between 50% nominal output torque level (277 in-lbs) and 300% nominal output torque level (1665 in-lbs) which lasted 19.3 hours, 151 data files were collected. This run was periodically stopped to allow oil samples and gear photos to be taken. When bringing the run to a stop for these purposes, data was sampled when the load dropped to 250%, 200%, 150%, 100% and 50%.

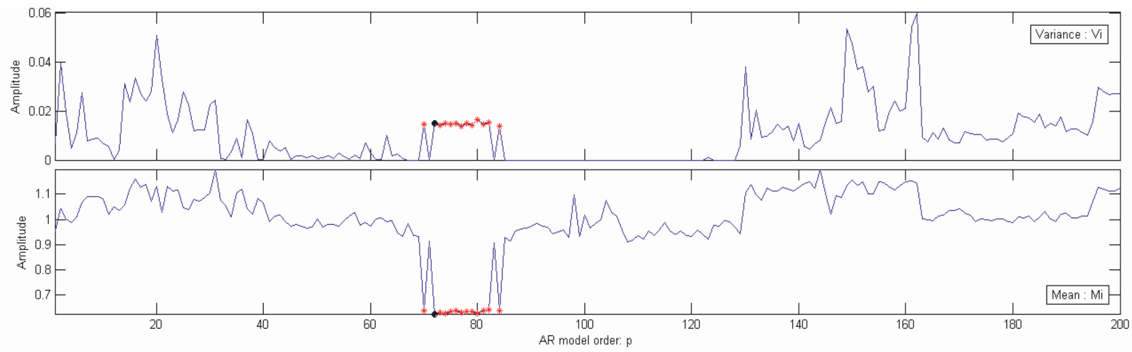
Figure 4(a) shows the load condition of TR#14. The input speed slightly varies between 1750 and 1752 rpm throughout the entire test-run. The gear was driven until there were seven broken teeth on the drive gear, as shown in Figure 4(b). A TSA signal of 2284 data point length at each inspection was obtained after an averaging operation over 87 ensembles. The GMR signals generated in this test-run were obtained by removing 12 sidebands on either side of every meshing frequency from the FFT spectrum of each TSA signal.

4.2 Condition detection

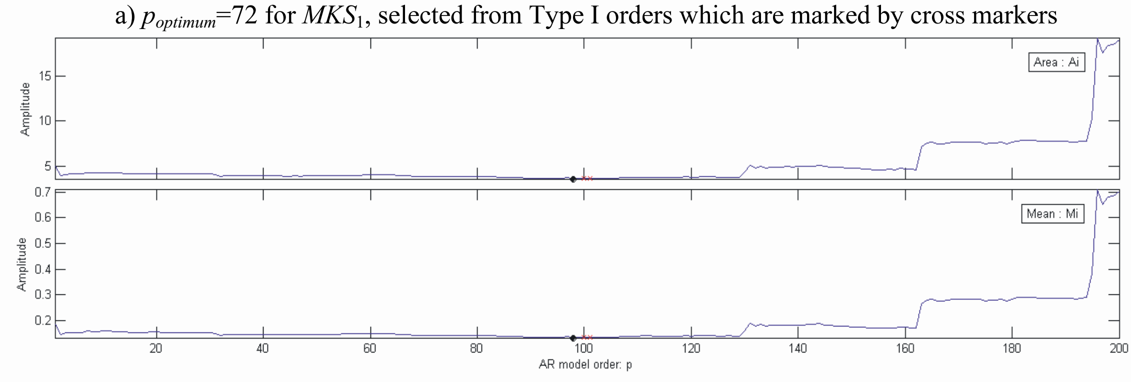
It is noteworthy that two previous studies conducted analysis of the same MDTB data sets and proposed two condition indicators termed the fault growth parameter (FGP) [15] and the modified fault growth parameter (FGP1) [16] for detection of gearbox deterioration, respectively. The comparative analysis will incorporate these two condition indicators.

Twelve GMR signals, i.e. [173, 174, 188, 189, 202, 203, 216, 217, 230, 231, 244, 245], under the 50% nominal output torque level and fifteen GMR signals, i.e. [180, 181, 182, 183, 194, 195, 197, 208, 209, 210, 211, 222, 223, 224, 225], under the 300% nominal output torque level are selected to form the control group GMR signals to establish the compromised AR model fitting on the healthy GMR signals under varying load conditions. Figure 5 shows the AR model order selection for MKS_1 and MKT_1. The selected optimum orders are 72 for MKS_1 and 98 for MKT_1.

Figure 6 shows the gear condition detection using MKS_1 and MKT_1 under the optimum AR model orders found in Figure 5 for as well as FGP and FGP1. Table 3 summarizes the performance evaluation of these four condition indicators for TR#14. It is found from Figures 6(c) and 6(d) that FGP and FGP1 are dependent on the sinusoidal load condition beyond the GMR signal #173. Further inspection indicates that the variations of FGP and FGP1 are opposite to that of load level at a few inspections. For example, the local peaks of FGP and FGP1 appearing at the 173rd, 202nd, and 216th inspections as indicated by circle marks



a) $p_{optimum}$=72 for MKS_1, selected from Type I orders which are marked by cross markers



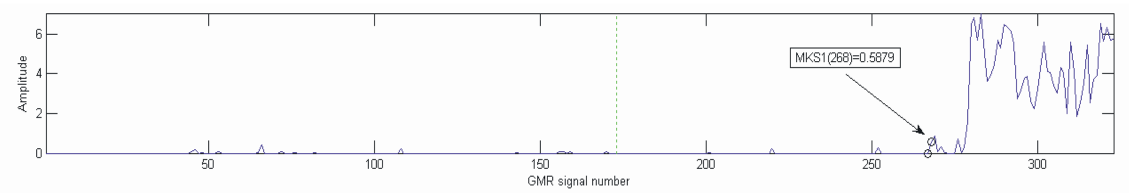
b) $p_{optimum}$=98 for MKT_1, selected from Type II orders, where the orders marked by cross markers are the orders with means of K-S outliers which are statistically equal to that of the $p_{optimum}$ via the Satterthwaite's t'-test at the significance level of 5%

Figure 5. AR model order selection, where the order marked by a dot in each subplot is the selected order $p_{optimum}$

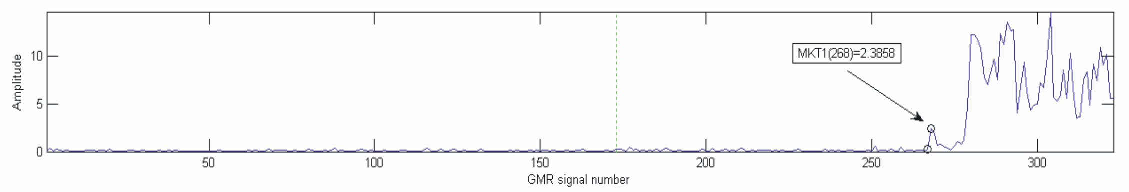
in Figures 6(c) and 6(d) respectively correspond to inspections carried out at 50% nominal output torque level. Such an inverse relationship may be attributed to the fact that, when the gearbox is in a healthy condition, major fluctuations are usually present in lightly loaded conditions, while heavily loaded conditions usually do not exhibit large instantaneous variations. In contrast, the proposed MKS_1 and MKT_1 do not present load-dependent behavior under the sinusoidal load condition as shown in Figures 6(a) and 6(b) respectively, where the normality assumption of AR model residuals is generally guaranteed at most inspections during the healthy period of this test-run.

Based on visual inspection of the traces of the four condition indicators, the occurrence of incipient gear fault may appear at the 268th inspection. For the purpose of reference, the amplitude of each condition indicator at the 267th inspection is also denoted by a circle mark in each subplot of Figure 6. It should be noted that the proposed MKS_1 is against the normality assumption at very few inspections during the healthy period of the investigated gear, i.e. the GMR signal interval [1:267] as shown in Figure 6(a). Due to distinct theoretical basis, the kurtosis based MKT_1 does not possess such a property, where the normality in the healthy condition of the gear cannot be guaranteed at all inspections as shown in Figure 6(b). Although the MKT_1 presents a stable trace within the GMR signal interval [1:267], our extensive tests that are not presented here have shown that MKT_1 is prone to more considerable variability than MKS_1. Such a disadvantage can be evidenced by examining the difference between MKS_1 and MKT_1 in Figures 3(a) and 3(b) or, alternatively, MKS_2 and MKT_2 in Figures 3(c) and 3(d).

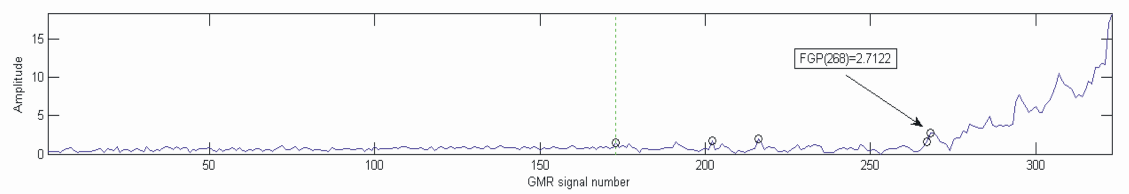
On the other hand, since FGP and FGP1 do not have a critical value to define the healthy condition of a gear, they are in need of a longer period to gain an understanding of around which numerical value FGP or FGP1 corresponds to the healthy condition of the investigated gear, where thus a prompt condition analysis cannot be carried out. It should be emphasized that the strong load-dependence of FGP and FGP1 to a great degree exacerbates such a disadvantage since, at every specific load condition, an individual examination is compulsory to find the numerical value of FGP and FGP1 which corresponds to the healthy condition of the investigated gear. Otherwise, the load induced jump of FGP or FGP1 may extremely possibly result in frequent false alerts and in turn incur maintenance cost in vain.



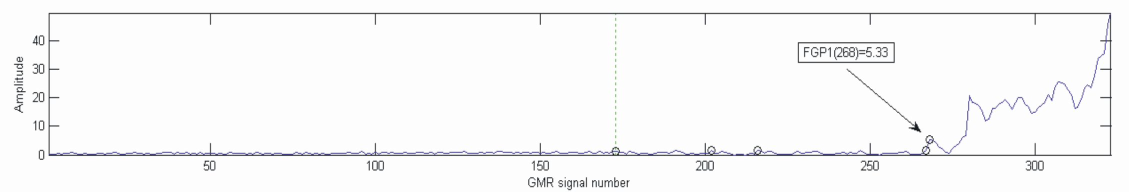
a) MKS_1 using AR(72) where the first alert is generated at the 268th inspection



b) MKT_1 using AR(98) where the first alert is generated at the 268th inspection



c) *FGP* where the first alert is generated at the 268th inspection



d) *FGP*1 where the first alert is generated at the 268th inspection

Figure 6. Gear condition detection of TR#14. Each GMR signal on the abscissa represents an inspection. The dotted line in each subplot denotes the GMR signal #173, i.e. the first inspection carried out under the sinusoidal load condition. The annotated inspections in each subplot correspond to the first alert for incipient fault and the false alerts.

Experimental analysis using alternative MDTB test-runs was also carried out. The proposed MKS_1 constantly demonstrates remarkable performance in comparison to a variety of condition indicators proposed in some recent studies. Therefore, the proposed MKS_1 significantly outperforms its counterparts proposed in relevant studies.

Table 3

Performance assessment of condition indicators for TR#14

Condition indicator	Load-independence	First alert for incipient fault
MKS_1	Yes	268
MKT_1	Yes	268
FGP	No	268
$FGP1$	No	268

5 CONCLUSIONS

In this study, gear vibration signals are divided into two categories: the control group GMR signals and the treatment group GMR signals. The control group consists of healthy condition GMR signals for modeling purposes, while the treatment group consists of regular lifetime GMR signals for validation and detection. This study proposed a novel compromised AR model fitting technique with the aid of hypothesis tests, which establishes a compromised AR model on the control group GMR signals. By compromised, we mean that the AR model of each healthy condition GMR signal in the control group has identical model order and model coefficients.

The model order is classified into two categories, i.e. Type I and Type II, and selected by means of two hypothesis tests, i.e. Bartlett's test and Satterthwaite's t'-test, respectively. The role of the Bartlett's test is to check, within all Type I orders, if the variance of K-S test statistics under the order with the least mean of K-S test statistics is statistically equal to the variance of K-S test statistics under the order whose variance of K-S test statistics is the least. The role of the Satterthwaite's t'-test is to check, within all Type II orders, if the order with the least mean of K-S outliers has alternative orders with statistically equal means of K-S outliers.

The calculation of model coefficients depends on the phase information being available or not. In consideration of the healthy condition GMR signals that were used to construct the compromised AR model not being in phase with the future-state GMR signals, an averaged coefficient vector is used.

Simulation and experimental analyses confirm the remarkable effectiveness of the proposed compromised AR model fitting technique in selecting the optimal order for gear condition detection under varying load conditions. In particular, the proposed K-S statistic based condition indicator MKS_1 demonstrates an appealing robustness to varying load conditions.

Furthermore, it is noteworthy that the gear condition detection technique proposed in this study is completely automatic in both modeling and detection. In particular, the proposed AR model order selection technique can be carried out automatically and is in need of no expert human supervision. In comparison with a previous study by Zhan et al. [12], the AR model order selection is greatly improved.

6 ACKNOWLEDGEMENTS

The authors are most grateful for the financial support provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada and Syncrude Canada Limited. The authors' gratitude also goes to the Condition-Based Maintenance Department in the Applied Research Laboratory at the Pennsylvania State University for providing the MDTB data.

7 REFERENCES

[1] McFadden PD. (1986) Detecting fatigue cracks in gears by amplitude and phase modulation of the meshing vibration. *ASME Transactions Journal of Vibration, Acoustics, Stress and Reliability in Design*, 108(2), 165–170.

[2] McFadden PD. (1988) Determining the location of a fatigue crack in a gear from the phase of the change in the meshing vibration. *Mechanical Systems and Signal Processing*, 2(4), 403–409.

[3] Dalpiaz G, Rivola A and Rubini R. (2000) Effectiveness and sensitivity of vibration processing techniques for local fault detection in gears. *Mechanical Systems and Signal Processing*, 14(3), 387–412.

[4] Wang W. (2001) Early detection of gear tooth cracking using the resonance demodulation technique. *Mechanical Systems and Signal Processing*, 15(5), 887–903.

[5] Sung CK, Tai HM and Chen CW. (2000) Locating defects of a gear system by the technique of wavelet transform. *Mechanism and Machine Theory*, 35(8), 1169–1182.

[6] Ohue Y, Yoshida A and Seki M. (2004) Application of the wavelet transform to health monitoring and evaluation of dynamic characteristics in gear sets. *Proceedings of the I MECH E Part J Journal of Engineering Tribology*, 218(1), 1–11.

[7] Nikolaou NG and Antoniadis IA. (2002) Rolling element bearing fault diagnosis using wavelet packets. *NDT & E International*, 35(3), 197–205.

[8] Andrade FA, Esat II and Badi MNM. (2001) Gear condition monitoring by a new application of the Kolmogorov-Smirnov test. *Proceedings of the I MECH E Part C Journal of Mechanical Engineering Science*, 215(6), 653–661.

[9] Kar C and Mohanty AR. (2004) Application of KS test in ball bearing fault diagnosis. *Journal of Sound and Vibration*, 269(1-2), 439–454.

[10] Hall LD and Mba D. (2004) Acoustic emissions diagnosis of rotor-stator rubs using the KS statistic. *Mechanical Systems and Signal Processing*, 18(4), 849–868.

[11] Wang W and Wong AK. (2002) Autoregressive model-based gear fault diagnosis. *Journal of Vibration and Acoustics*, 124(2), 172–179.

[12] Zhan YM, Makis V and Jardine AKS. (2006) Adaptive state detection of gearboxes under varying load conditions based on parametric modeling. *Mechanical Systems and Signal Processing*, 20(1), 188–221.

[13] MDTB test-run data CD: TR#14, Condition-Based Maintenance Department, Applied Research Laboratory, The Pennsylvania State University.

[14] Byington CS and Kozlowski JD. (1998) Transitional data for estimation of gearbox remaining useful life. Applied Research Laboratory, The Pennsylvania State University, MDTB test-run data CDs.

[15] Miller AJ. (1999) A new wavelet basis for the decomposition of gear motion error signals and its application to gearbox diagnostics. M.S. thesis, The Graduate School of The Pennsylvania State University.

[16] Lin D, Wiseman M, Banjevic D and Jardine AKS. (2004) An approach to signal processing and condition-based maintenance for gearboxes subject to tooth failure. *Mechanical Systems and Signal Processing*, 18(5), 993–1007.

[17] Law AM and Kelton WD. (1991) Simulation Modeling and Analysis. McGraw-Hill, New York, 387–390.

7.1 APPENDIX A: Time-Varying AR Model Based on a Noise-Adaptive Kalman Filter

An AR process is a discrete-time multivariate linear stochastic process given by

$$y_i = \sum_{k=1}^{p} A_k y_{i-k} + \varepsilon_i \tag{A.1}$$

for i = 1, 2, …, N, where N is the sample size, p the order of VAR model, y_i the ith measurement vector of dimension $d\times1$, A_k the kth $d\times d$ coefficient matrix of the measurement y_{i-k}, and ε_i an $d\times1$ sequence of zero-mean white Gaussian measurement noise. Considering the non-stationary cases, the coefficient matrices of the AR model are assumed time-varying

$$y_i = \sum_{k=1}^{p} A_k(i) y_{i-k} + \varepsilon_i \qquad \text{(A.2)}$$

The time-varying model coefficients are in need of an adaptive estimation algorithm. The Kalman filter is employed here. With the following notation:

$$a_i = vec([A_1(i), A_2(i) \ldots, A_p(i)]^T) \qquad \text{(A.3)}$$

$$Y_i = (y_i^T, y_{i-1}^T, \ldots, y_{i-p+1}^T) \qquad \text{(A.4)}$$

$$C_i = I_d \otimes Y_i^T \qquad \text{(A.5)}$$

where I_d is an $d \times d$ identity matrix, $\otimes$ is the Kronecker product, an appropriate state space representation of the VAR model with stochastic coefficients can be given by

$$a_{i+1} = f_i(a_i) + \nu_i \qquad \text{(A.6)}$$

$$y_i = C_{i-1}^T a_i + \varepsilon_i \qquad \text{(A.7)}$$

where a_i is the $pd^2 \times 1$ state vector, ν_i an $pd^2 \times 1$ sequence of zero-mean white Gaussian state noise, C_{i-1}^T of dimension $d \times pd^2$, and the noise covariance matrices:

$$E\{\nu_k \nu_i^T\} = Q_k \delta_{k-i} \qquad \text{(A.8)}$$

$$E\{\varepsilon_k \varepsilon_i^T\} = R_k \delta_{k-i} \qquad \text{(A.9)}$$

where T denotes transposition, δ denotes the Kronecker delta sequence, Q_k denotes the covariance matrix of ν_i and R_k denotes the covariance matrix of ε_i. Suppose that the evolution law of the state vector a_i is a random walk process which results in a state space representation

$$a_{i+1} = a_i + \nu_i \qquad \text{(A.10)}$$

$$y_i = C_{i-1}^T a_i + \varepsilon_i \qquad \text{(A.11)}$$

Thus, with the aid of a standard Kalman filter the following recursive prediction equations

$$P_{i|i-1} = P_{i-1|i-1} + Q_{i-1} \qquad \text{(A.12)}$$

$$\hat{a}_{i|i-1} = \hat{a}_{i-1} \qquad \text{(A.13)}$$

where $P_{i|i-1}$ is the one-step ahead prediction of the state covariance matrix, $P_{i-1|i-1}$ is the estimate (error) covariance matrix of the state vector, $\hat{a}_{i|i-1}$ is the one-step ahead predication of the state vector, $\hat{a}_{i-1}$ is the optimal filtering estimate of the state vector a_{i-1} and updating equations

$$G_i = P_{i|i-1} C_{i-1} [C_{i-1}^T P_{i|i-1} C_{i-1} + \hat{R}_i]^{-1} \tag{A.14}$$

$$P_{i|i} = [I - G_i C_{i-1}^T] P_{i|i-1} \tag{A.15}$$

$$\begin{aligned} \hat{a}_{i|i} &:= \hat{a}_i \\ &= \hat{a}_{i|i-1} + G_i (y_i - C_{i-1}^T \hat{a}_{i|i-1}) \end{aligned} \tag{A.16}$$

can be obtained for $i = 1, 2, \ldots, N$, where G_i is the Kalman gain and the covariance matrix Q_i of ν_i is assumed to be known *a priori*. The innovations sequence is defined as

$$z_i := y_i - C_{i-1}^T \hat{a}_{i|i-1}, \tag{A.17}$$

which leads to the estimate of R_i, using the i most recent residuals, given by

$$\hat{R}_i = \frac{1}{i-1} \sum_{k=1}^{i} (z_k - \bar{z})(z_k - \bar{z})^T, \tag{A.18}$$

where

$$\bar{z} = \frac{1}{i} \sum_{k=1}^{i} z_k . \tag{A.19}$$

is the mean of the innovations up to the ith time instant. Therefore, an adaptive Kalman filter for estimating the state as well as the noise covariance matrices is termed the noise-adaptive Kalman filter (NAKF).

7.2 APPENDIX B: Kolmogorov-Smirnov Goodness-of-Fit Test

The Kolmogorov-Smirnov (K-S) goodness-of-fit test compares an empirical distribution function with the distribution function of the hypothesized distribution [17]. To define the K-S statistic, we must first define an empirical distribution function. For the K-S test, we define an empirical distribution function $F_n(x)$ from our data $X_1, X_2, \ldots, X_n$ as

$$F_n(x) = (\text{number of } X_i\text{'s} \le x) / n \tag{B.1}$$

for all real numbers x. Thus, $F_n(x)$ is a (right-continuous) step function such that $F_n(x_{(i)}) = i/n$ for $i=1,2,\ldots,n$. If $\hat{F}(x)$ is the fitted distribution function, a natural assessment of goodness of fit is some kind of measure of the closeness between the functions F_n and $\hat{F}$. The K-S test statistic D_n is simply the largest (vertical) distance between $F_n(x)$ and $\hat{F}(x)$ for all values of x and is defined formally by

$$D_n = \sup_x \{|F_n(x) - \hat{F}(x)|\} \tag{B.2}$$

[The 'sup' of a set of numbers A is the smallest value that is greater than or equal to all members of A. The 'sup' is used here instead of the more familiar 'max' since, in some cases, the maximum may not exist. For example, if $A = (0,1)$, there is no maximum but the 'sup' is 1.] D_n can be computed by calculating

$$D_n^+ = \max_{1 \le i \le n} \left\{ \frac{i}{n} - \hat{F}(x_{(i)}) \right\}, \quad D_n^- = \max_{1 \le i \le n} \left\{ \hat{F}(x_{(i)}) - \frac{i-1}{n} \right\} \tag{B.3}$$

and finally letting

$$D_n = \max \left\{ D_n^+, D_n^- \right\} \tag{B.4}$$

Suppose that the hypothesized distribution is N(μ, σ^2) with both μ and σ^2 unknown. We can estimate μ and σ^2 by $\overline{X}(n)$ and $S^2(n)$, respectively, and define the distribution function to be that of the N($\overline{X}(n)$, $S^2(n)$) distribution; i.e., let $\hat{F}(x) = \Phi\left\{ [x - \overline{X}(n)] / \sqrt{S^2(n)} \right\}$, where Φ is the distribution function of the standard normal distribution. Using this $\hat{F}$ (which has estimated parameters), D_n is computed in the same way, but different critical points must be used. An accurate approximation, which obviates the need for large tables, is provided by Law and Kelton [17]; namely, we reject H_o if

$$\left(\sqrt{n} - 0.01 + \frac{0.85}{\sqrt{n}} \right) D_n > c'_{1-\alpha} \tag{B.5}$$

where α is always 0.05 in this study and thus $c'_{1-\alpha}$ takes the value of 0.8950 as shown in the Table 6.14 on page 390 by Law and Kelton [17].

DEVELOPMENT OF AN OPTIMUM VOLTAGE STEP CONTROL STRATEGY FOR A ROBOTIC MANUFACTURING PROCESS

CK Mechefske and C Zheng

Department of Mechanical and Materials Engineering, Queen's University, Kingston, Ontario, Canada K7L 3N6

Abstract: This paper addresses the development of optimum current step settings used at a local manufacturing plant. The step function and the effects on other operation process parameters are obtained through reverse and correlation analysis of Target Parameter C (for confidential concerns, all related information are replaced by close item as required), Parameter B and real one. The Secondary Current for any operation sequence is then obtained based on this control methodology. The correlation between operation order and tip wear value are also determined by modeling the real data. A simulation model for optimum step control for a robotic manufacturing system has been developed. Based on some assumptions, simulations are carried out to control two important indicators within their proper limits. One is the Parameter E consisting of real secondary current and Parameter D, which should be constant or vary only slightly through the whole process. The other is current density. Each step can be adjusted and optimized and the indicator trend curves showing operation sequences per step and expected total operation sequences will be displayed. Results show that the step setting can be optimized to control the operation or step current to get good quality. At the beginning of each cycle, Real Secondary Current can be set very close to the Target one(which is much lower than the existing setting level) , so power is minimized and the tip useful life is extend.

Key Words: Robotics Spot Welding, Optimum Current Control, Test Data Modelling, Simulation

1 INTRODUCTION

Robotic manufacturing process is becoming increasingly important for the modern industry due to its high efficiency, reliability and productivity. It is well known that it is the generated heat (shown in equation 1) that produces the kind of process.

$$\text{Heat} = I^2Rt \tag{1}$$

Where, I is the real secondary current, R is the total resistance and t s the operation cycle.

As the number of operation increase, electrode tip wear and the contact area between the tip and the workpiece becomes larger resulting in lower current density. As the current density gradually decreases there is increased risk of failure. Therefore, a current controller (stepper booster) is required to guarantee that enough heat will be generated during the whole process. This kind of stepper booster (most linearly stepper) is used to regulate the amount of current which is added to the base heat that starts out with new tips. It has the following functions: 1) It automatically steps current to compensate for electrode wear and contamination ensuring better, more consistent operation. 2) It counts the number of operation made.3) increases the same current in each discrete steps. 4) It enables the operation gun to maintain a constant current density or within acceptable control limit[1].

This stepper booster is expected to be fine-tuned to match the actual wear and contact area change that process causes on the tips at all stages. In reality, it is not easy to achieve such an objective because the process mechanism is so complex. The timing of step changes and the amplitude of the current to be used in each step is not well understood.

There is a tradeoff between too much current, resulting in expulsion and too little current, resulting in under-size nugget. Any change in the schedule, such as operation sequence force or heat time, will affect the wear and generated heat on the tips as well. Therefore, many companies and researchers have involved in the research of stepper control. For example: WTC, Atek Corperation, Diamlerchrysler, Toshiba and etc. [2-4] have developed many useful heat control tools for industries and they can be set to use in different applications. Jou Min[5] studied the real time monitoring of this process for the fabrication of sheet metal assemblies, and explored the phenomenon of how signal indicative (electrode displacement) of strength and quality changes due to the input (percentage heat input) for various sheet steels used in industry. I. Komine et.al.[6] studied the heat control of this process in steel pipe production. All these have very successful applications in reality. However, the stepper booster does not track the change in surface area. Even though the current is regulated, the current density is overlooked. Unfortunately, inadequate current density usually produces non intact process. Therefore, it is also crucial to have a proper current density in the area where the process is to be made. This will require an active adjustment of the stepper boost which has become a real and practical problem for some manufacturing plants.

In this paper, the current step control strategy in a robotic manufacturing process will be studied. Our industry partner had invested more than $2M to realize online quality and process control which has built up solid foundation for step optimization. Based on field data models, we use simulation to obtain the optimum operation sequence strategy for each step (the appropriate amount of current increment and when to transfer to the next step). The final aim is not only to get good product every time, but also to extend the tip life as long as possible and minimize the energy usage.

2 FIELD DATA ANALYSIS AND MODELING

The experimental work has been conducted on the robotic manufacturing factory which has significant experience on step optimization in its real production line. Both the 5 step and 3 step methodologies have been used and each of them has its successful application. But there is still room for optimization because the current steps used now mostly come from experience and may not necessarily be the best.

A monitoring system has been employed to realize online quality and process control. Necessary information is automatically collected and input into the PLC, and then transferred and stored in a database. Data of 21 tip shifts from a servo gun of robot LHsub1 between June 1 and June 24 in 2003 have been obtained. Data sample of tip 18 is shown in Table 1. The operation robot performs a tip check and tip wear compensation with no current passing through after operating several sequences, which is good for tip wear compensation and heat dissipation.

Table 1: Data sample for tip 18

Log #	Part #	Operation Sequence #	Tip Wear	Parameter A	Parameter B	Target Parameter C	Parameter D	Parameter E	Stepper	Note
4761	210	1	65519	568	124	12500	67	195	1	
4762	210	2	65519	572	123	12500	69	190	1	
4763	210	3	65519	572	123	12500	67	194	1	
4764	210	4	65519	570	123	12500	69	189	1	
4765	210	24	65519	0	0	0	0	0	1	Tip Check
4766	211	1	65515	575	124	12500	67	194	1	
4767	211	2	65515	574	123	12500	69	188	1	

2.1 Step Setting

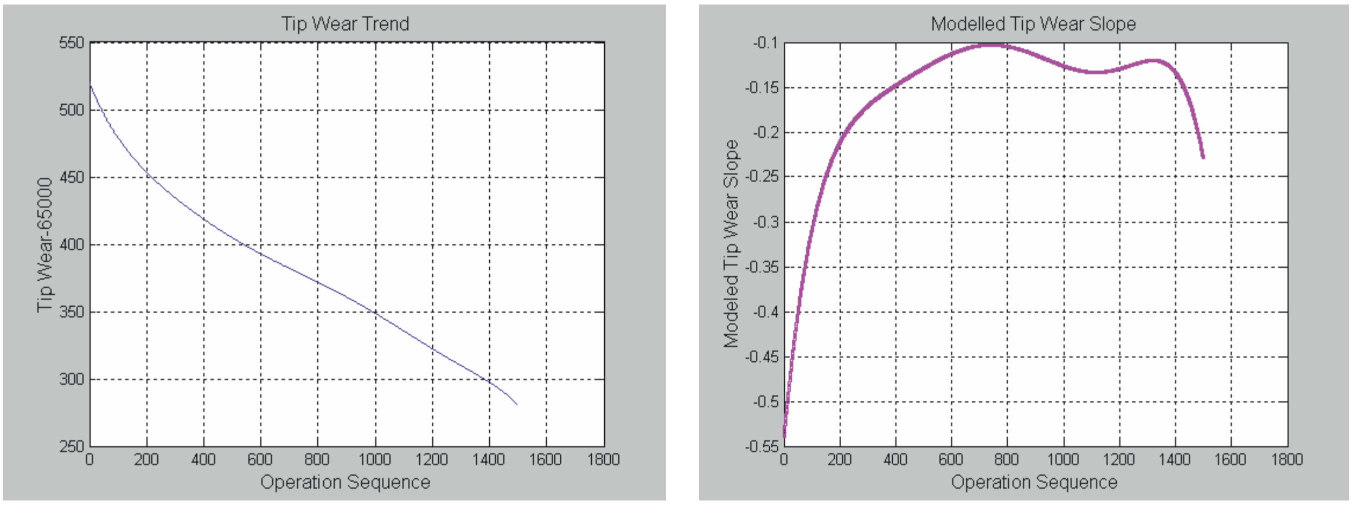
In this experiment, a 3 step methodology is employed. The step setting for some tip shifts is shown in Table 2. This table indicates that step 1 ends around 100 operation sequences, step 2 ends at 400 operation sequences and step 3 is used for the remainder.

Table2: Step Number Distribution

Tip No.	Total Operation number	Step=1	Step=2	Step= 3
4	1448	100	400	948
5	1428	96	400	932
6	1464	100	400	964
11	1444	100	400	944
12	1448	100	400	948

2.2 Tip Wear Data Analysis and Modeling

The repeated use of a set of tips, heat and force tend to increase the effective radius of the tips (mushrooming), and flatten the contact area. The average model and its slope trend are shown in Figure 1[7]. From tip wear results, it is clear that the tip wear decreases sharply before 200 operation sequences. Therefore, if an earlier step setting is adopted, tip life may be extended.



a) Tip wear general model **b) Slope of tip wear model**

Figure 1: Tip wear model and its slope trend

2.3 Parameter D and A Data Analysis and Modeling

Parameter D is the actual % heat fired. It follows a similar increasing trend shown in Figure 2, but drops again at the end of each tip life. The heat must be taken back out (be reset to the 1st stepper level) because new tips will be installed, otherwise it will result in expulsion and tip sticking after a tip change. This Parameter A average of all tips varies slightly, as shown in Figure 3.

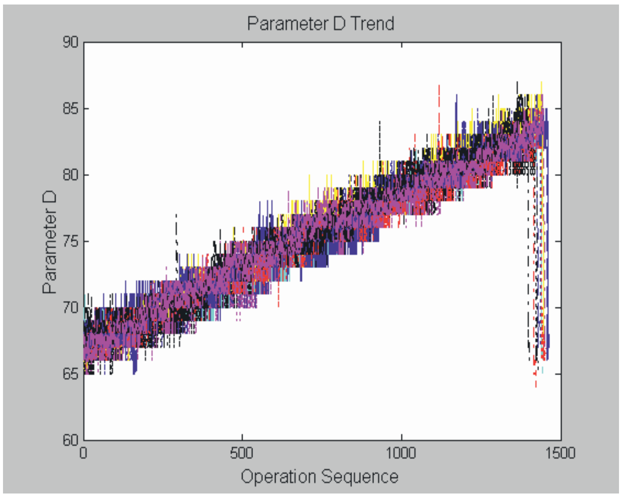


Figure 2: Parameter D trend

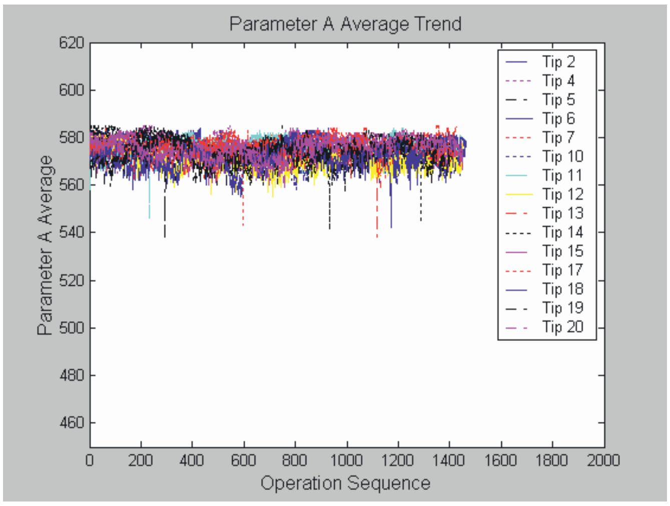


Figure 3 Parameter A Average trend

2.4 Parameter E Data Analysis and Modeling

Measured parameter E.

The Parameter E represents one percent of current increases or equals the secondary current divided by Parameter D. Parameter E should be constant or vary only very slightly through the whole process. The step control can compute the output current and the percentage that it fired after each cycle pulse.

$$E = \frac{I_{\sec}}{D} \tag{2}$$

Where I_{sec} is the real secondary current.

Normalized parameter E.

The above Parameter E equation assumes that the operation bus is constant. However, current flow is not only affected by the resistance of the tools but also by the Parameter A of the bus. As the number of operation increases, Parameter A typically increases while current decreases. Resistance at the workpiece is not constant. So the Normalized Parameter E, based on Parameter A average, is introduced in order to strictly observe the resistance variable and is defined as follows [8]:

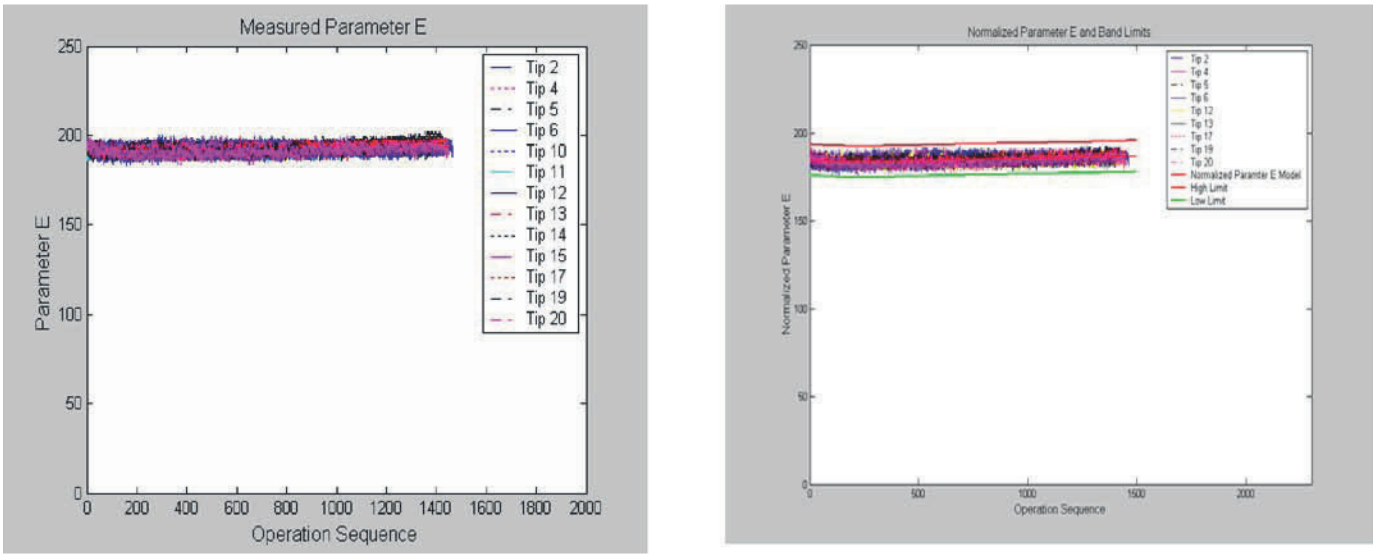
$$E_{normalized} = \frac{E_{absolute} * A_{norminal}}{A_{actual}} \tag{3}$$

The measured and Normalized Parameter E general models representing the trends of all tips are as follows and shown in Figure 4.

$$E(z) = -0.00061791x^3 + 0.031178x^2 - 0.15953x + 191.73 \tag{4}$$

$$\text{Normalized\_E(z)} = -4.4813e-005x^5 + 0.0022658x^4 - 0.042418x^3 + 0.36065x^2 - 1.1194x + 184.69 \tag{5}$$

Where, z=1,2,3,…1500 and x=z/100.



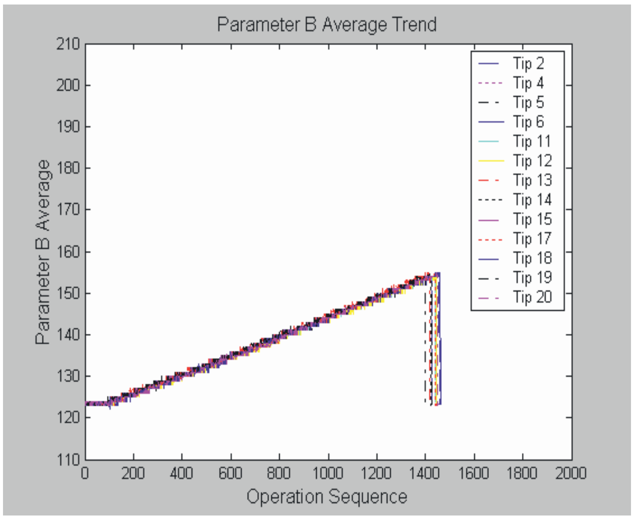
a) Measured Parameter E **b) Normalized Parameter E Model and Band Limits**

Figure 4: Parameter E trend and models

2.5 Parameter Correlation Analysis

Parameter B.

The measured Parameter B includes a measure of the maximum (high), the average (average) and the minimum (low). Because the stepper controller has the ability to control the current very accurately, most of the operation sequences do not vary significantly from the average. Among these three, Parameter B Average of all these tips follows a well behaved gradually increasing trend shown in Figure 5. It is nearly constant at step 1, then increases continuously (almost linearly) during steps 2 and 3.



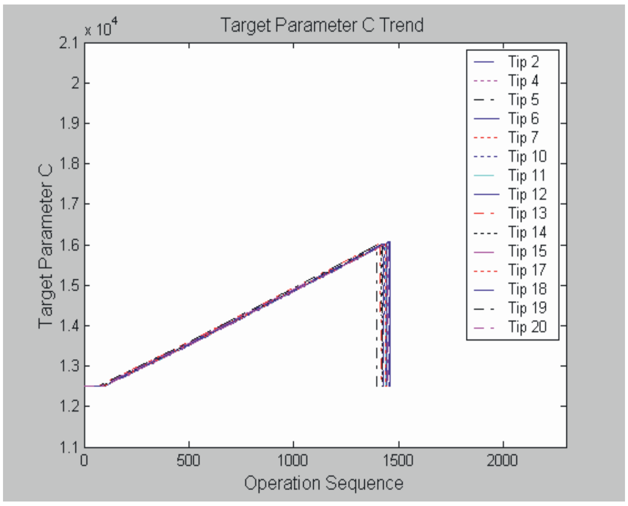


Figure 5: Parameter B Average Trends   Figure 6: Target Parameter C Trends

Target Parameter C.

Since the step controller is mainly designed to increase the current, the information of how the current step setting affects the current increment can be obtained by analyzing it. The Target Parameter C value is pre-programmed according to the tip real situation, which has the very similar trend as that of Parameter B and is shown in Figure 6. It relates to the step setting and is therefore valuable to analyze carefully in order to get more details about its function. In this paper, Target Parameter C Slope is obtained and the result for each sequence of Tip#20 is shown as Figure 7. From the operation sequence Target Parameter C trends, it is clear that the increment is zero for step 1 and 10 Amps for nearly each sequence at step 2. The Target Parameter C of step 3 alternatively increases by 11 Amps for 2 continuous operation sequences and then 10 Amps for 1 operation sequence(averagely 10.667Amps). Table 3 shows the results of current steps 2 and 3 and the effect on Target Parameter C.

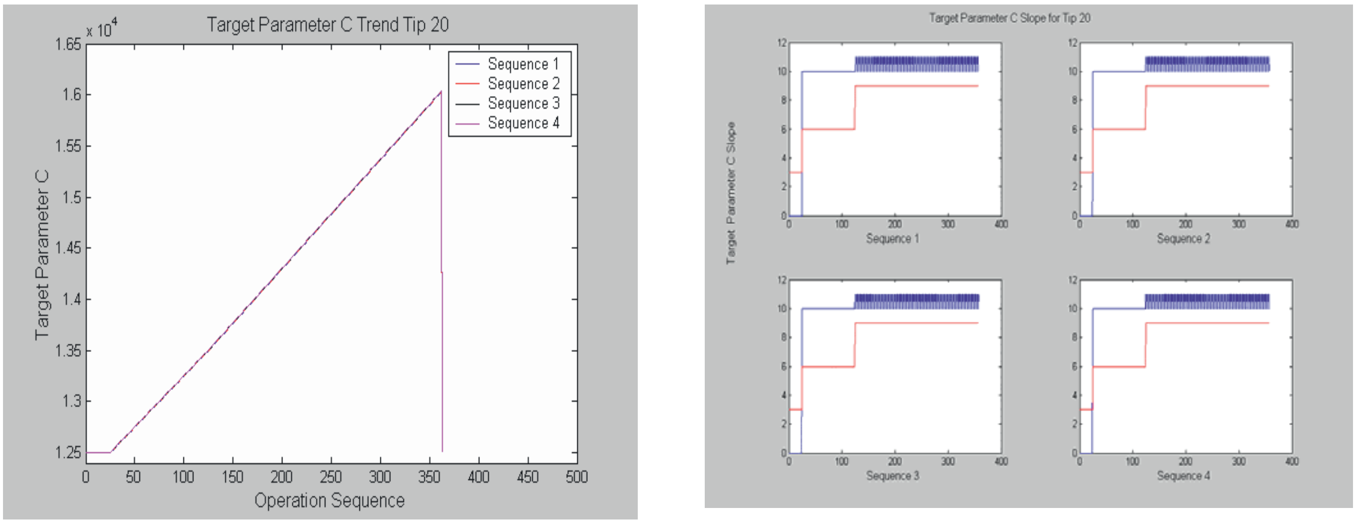


Figure 7: Target Parameter C and its Slope Behavior for each Sequence of Tip 20

The Target Parameter C for each Operation sequence starts at 12500 Amps. At the 25th part (100 operation sequences), step 2 begins. The Target Parameter C for operation sequence 1, 2, 3 and 4 starts at 12500, 12502, 12505 and 12507 respectively with the same increment as used in step 2. Therefore, the total current increment can be obtained. For Tip#20, the total operation sequence number is 1452, therefore there are 363 operation sequences in each of the four sequences. The total current increment is 3538.67 Amps and the increment percentage is 28.31% compared with the Target Parameter C of the first operation sequence(12500 Amps), which is listed in Table 4.

Correlation Analysis.

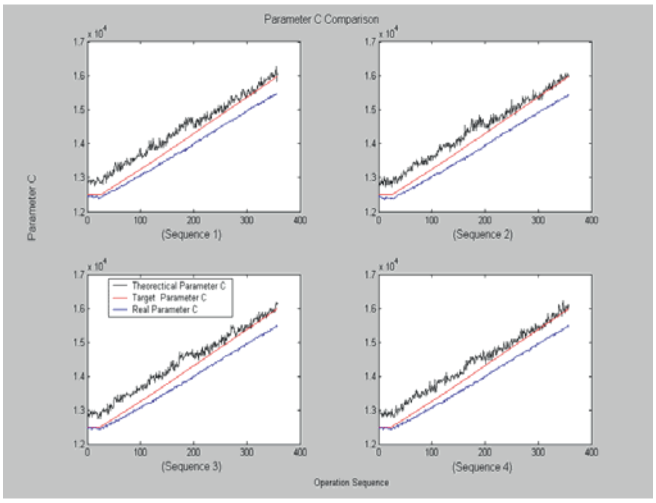
From Equation 2, it is clear that Real Secondary Current equals the Parameter E times Parameter D. Current Correlation among Real Secondary Current, Target Parameter C and Theoretical one(Parameter B times turn ratio) for each sequence of Tip#20 are shown in Figure 8. The current difference between the real and Target Parameter C of all tips is shown in Figure 9.

Table 3: Step effects for Target Parameter C

Tip	Step 2				Step 3			
	Operation Sequence 1	Operation Sequence 2	Operation Sequence 3	Operation Sequence 4		Operation Sequence 2	Operation Sequence 3	Operation Sequence 4
4	10	10	10	10.01	10.6652	10.6695	10.6652	10.6652
6	10	10	10	10.01	10.6667	10.6667	10.6667	10.6667
13	10	10	10	10.01	10.6638	10.6681	10.6681	10.6638
17	10	10	10	10.01	10.6638	10.6681	10.6681	10.6638
18	10	10	10	10.01	10.6639	10.6681	10.6681	10.6639
19	10	10	10	9.97	10.6638	10.6681	10.6681	10.6652
20	10	10	10	10.01	10.6667	10.6667	10.6667	10.6667

Table 4: Target Parameter C increment for tip 20

Tip	Total operation sequences	Sequence operation number	Sequence 1			Total current increment(Amps)	Increment Percentage
			Step 1	Step 2	Step 3		
20	1452	363	25	100	238	100*10+238*10.6667=3538.67	28.31%



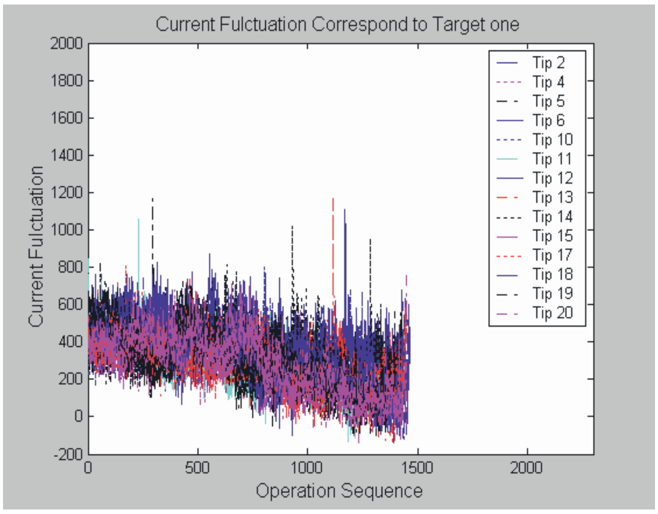


Figure 8: Parameter C Correlation **Figure 9. Parameter C Fluctuation for Tips**

From results, it is clear that, for quality concerns, both Real and Theoretical Parameter C have been programmed to be constant and much higher than the target one at step 1. However, this will greatly deteriorate the tip life, especially at an early age because of the small contact area and high pressure. It also wastes a lot of power on those operation sequences before the 100th operation sequence. At step 2, the real and Target Parameter C both increase, but with different rates. The difference between both of them becomes less as operation progresses. Finally, the Real Parameter C becomes equal to or less than its Target one at about 1000 or 1200 operation sequences. This demonstrates that the Real Parameter C could be a good indicator of operation sequence quality and Target Parameter C can be used as a reference. As the number of operation increases, resistance increases while operation parameter C decreases. This also verifies that the real one needs to be increased using the step settings to guarantee the current density is maintained within set limits with the progress of tip mushrooming.

3 STEP AND TIP WEAR SIMULATION

3.1 Simulation Assumptions

For simulation simplicity, the following assumptions have been made:

1) The same working conditions as exist on the real production line are adopted.

2) Parameter D follows the general models obtained.

3) Since current operation sequence quality is near 100% at this plant, we can use the 100th operation sequence current density as a standard to obtain the starting current point at the 1st operation sequence, based on the same current density.

4) Simulation will continue as long as control indicators fall within predefined acceptable limits

5) The same tip is used and tip wear development samples are shown in Figure 10 and 11. Table 5 shows the Tip contact area model characteristics. The tip contact area development is based on the Weibull Function and will be modified by later experimental work using the same equipment

Table 5: Electrode Dimension (in inches)[9]

Tip model	Operation Face A	Gage Line Diameter B	Length C	O.D.	Replacement Diameter D1	Area ratio A/A1	Contact Area Model
GCAP266	0.375	0.6225	1	0.750	13mm	0.5917	z=1-exp(-3.5*(t+0.4).^1.5)

Note: t=N/100 (N--Operation order)

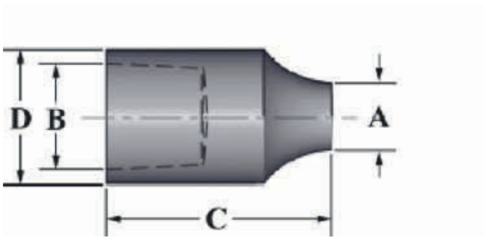


Figure 10: Tip Model

a) New tip **b) Medium tip wear** **c) Large tip wear**

Figure 11: Tip wear samples (Model MPG266 GCAP)

3.2 3.2 Simulation

Assume current at operation sequence order is z, then the proposed Current at any point z is obtained and listed in Table 6. The current trend is shown in Fig. 12.

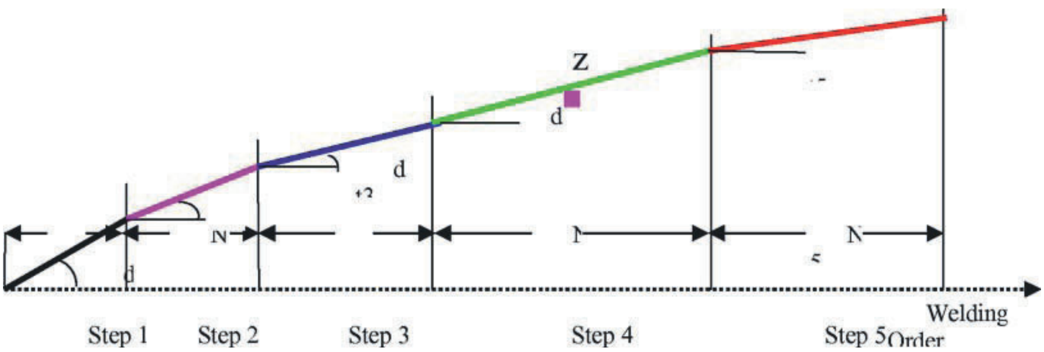


Figure 12: Proposed Current

The simulation interface is shown in Figure 13. This simulation is carried out to control two important indicators within their proper limits. One is the Parameter E consisting of Real Secondary Current and Parameter D, which should be constant or vary only slightly through the whole process (i.e. within 8%). The other is current density consisting of Real Secondary Current and Real Contact Area, which should fall in acceptable band limits during all stages (i.e. within 20%). The acceptable fluctuation percentage and each step current increment can be adjusted and optimized by the user until the final maximum operation sequence number is obtained. After simulation, the trends of current density, Parameter E and Real secondary Current will be displayed on the interface. Each stepper operation number and two indicator control limits are also shown on the interface. Each step can be adjusted and optimized and the indicator trend curves show operation amount per step and expected total operation amount will be displayed. Results show that the step setting can be optimized to give more good operation sequences throughout the life of the tip. At the beginning of each cycle, Real Secondary Current can be set very close to the Target (which is much lower than the existing setting level), so power is minimized and the tip useful life is extend.

Table 6: Proposed Current

Step	Current Increment	Operation number	Secondary Current at any point z	Starting Point of z
Step 1	dt1	N1	PC1=PC+z*dt1	0
Step 2	dt2	N2	PC2(z)=PC+N1*dt1+(z-N1)*dt2	N1+1
Step 3	dt3	N3	PC3(z)=PC+N1*dt1+N2*dt2+(z-N2-N1)*dt3	N1+N2+1
Step 4	dt4	N4	PC4(z)=PC+N1*dt1+N2*dt2+N3*dt3+(z-N3-N2-N1)*dt4	N1+N2+N3+1
Step 5	dt5	N5	PC5(z)=PC+N1*dt1+N2*dt2+N3*dt3+N4*dt4+(z-N4-N3-N2-N1)*dt5	N1+N2+N3+N4+1

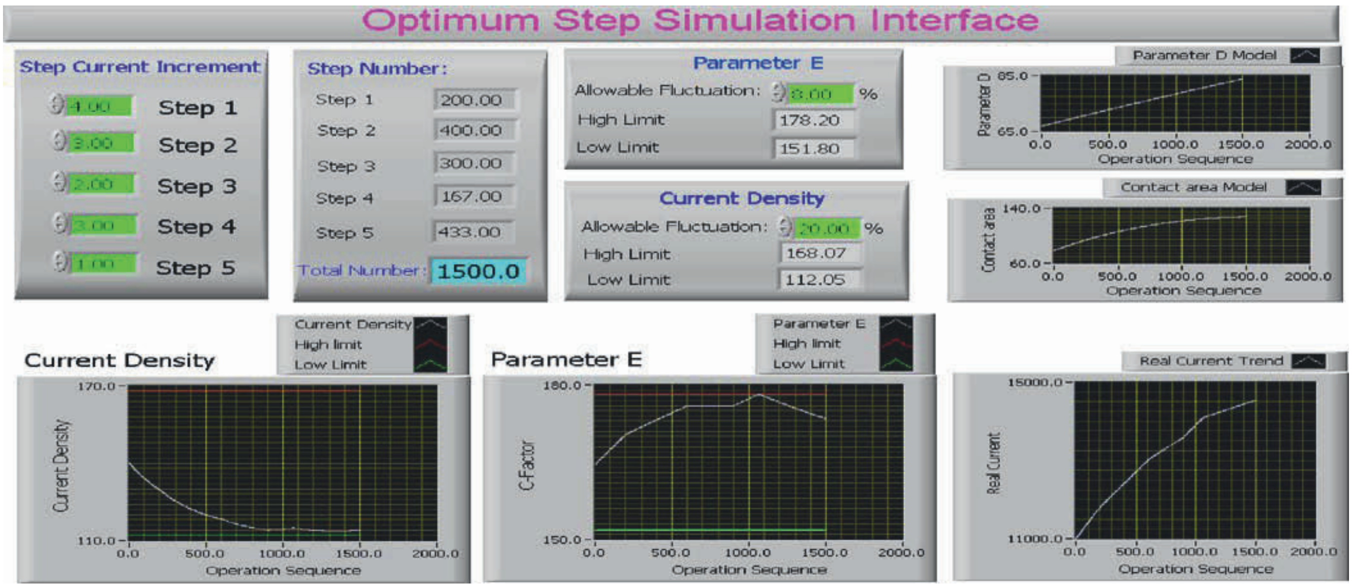


Figure 13 Simulation interface

4 SUMMARY

This paper addresses the current and tip wear problem for a robotic manufacturing process. The step compensation methodology is investigated. Then the development of optimum current step settings used at a local manufacturing plant are introduced, and then all the information related to step control from the real production line has been analyzed and modeled. The step function and the effects on other operation process parameters are obtained through reverse and correlation analysis of Target Parameter C, Theoretical and Real Secondary Current. The proposed Secondary Current for any operation sequence is then obtained based on this control methodology.

A simulation model for optimum step control for this kind of robotic manufacturing process has been developed. The correlation between tip contact area and operation count order is also proposed using a Weibull Function. Some assumptions are made and simulations are carried out to control two important indicators within their proper limits. One is the Parameter E consisting of Real Secondary Current and Parameter D, which should be constant or vary only slightly through the whole process. The other is current density consisting of Real Secondary Current and Real Contact Area, which should fall in an acceptable band during all stages. Each step can be adjusted and optimized and the indicator trend curves showing operation amount per step and expected total operation amount will be displayed. Results show that the step setting can be optimized to give more good operation sequences throughout the life of the tip, save energy and to extend the tip useful life. It is concluded that it is worthwhile to consider the Optimum step setting after using the information from this simulation.

5 ACKNOWLEDGEMENTS

Thanks are extended to industrial partner for providing the opportunity to work on this project. We further thank Dr. Kornel Farkas, Mr. Brian Morris for providing us with their help in the experimental work. We appreciate Dr. Sam Li and Mr. Barry Kavanaugh for their valuable information and suggestions for this project

6 REFERENCES

[1] Buchanan GH. (1989) Resistance Welding Manual, Forth Edition, Philadelphia, PA. USA, RWMA Publications

[2] http://www.medar.com

[3] http://www.atekcorp.com

[4] http://www.updatetechnology.com

[5] Min J. (2003) Real Time Monitoring Weld Quality of Resistance Spot Welding for the Fabrication of Sheet Metal Assemblies, Journal of Materials Processing Technology, Vol. 132, Issue: 1-3, pp. 102-113.

[6] Komine, I, Takahash I and Ishiro S. (1987) Heat Control for Electric Resistance Welding in Steel Pipe Production. Control Systems Magazine, IEEE, Vol. 7, Issue: 5, pp.10–14.

[7] Zeng C, Mechefske CK and Liaw J. (2005) A Simulation Model for Opportunistic Electrode Replacement in a Robotic Spot Welding System", International Journal of Computer Integrated Manufacturing, In press.

[8] Special Report - Normalized C-Factor, Engineering Department at Welding Technology Corporation.

[9] http://www.cmwinc.com, Spot WeldingData - Optimum Conditions Schedule for Spot Welding Low Carbon Steel - SAE1010

[10] http://www.spotweldingconsultants.com, Contacts Metals Welding.

OPPORTUNISTIC ELECTRODE REPLACEMENT IN A ROBOTIC SPOT WELDING SYSTEM

Mechefske CK and Zheng C

Department of Mechanical and Materials Engineering, Queen's University, Kingston, Ontario, Canada K7L 3N6

Abstract: This paper addresses the electrode replacement problem for a robotic resistance spot welding system consisting of k non-identical robots and the general maintenance optimization problem in a manufacturing setting. The electrode tip wear compensation methodology is introduced. Three kinds of replacement: corrective replacement, preventive replacement and opportunistic replacement policies are described. Polynomial models have been used to predict future tip behavior by modeling the tip wear from field data. Based on this, a tip wear trend as well as high and low limits for tip usage are established. Subsequently, the recommended and maximum number of welds left on a given tip and the corresponding remaining lifetime for a specific tip can be obtained. The expected failure time or replacement time can also be predicted. Through a careful cost investigation, a simulation model for electrode opportunistic replacement in a 6-robot resistance welding system has been developed. Based on some assumptions made, an optimum electrode opportunistic replacement interval achieving the maximum saving is acquired. The total expected cost, tip replace number, production line stoppage times and next replacement time can also be obtained.

Key Words: Opportunistic Maintenance, Resistance Spot Welding, Tip Replacement, Optimization Simulation.

1 INTRODUCTION

Spot resistance welding is the most inexpensive and the most common form of resistance welding. Robotic spot welding is becoming increasingly important for the automotive industry due to the fact that there are between 3,000 and 5,000 spot welds in a typical car. Consequently, the integrity of spot welds plays a critical role in the reliability of cars.

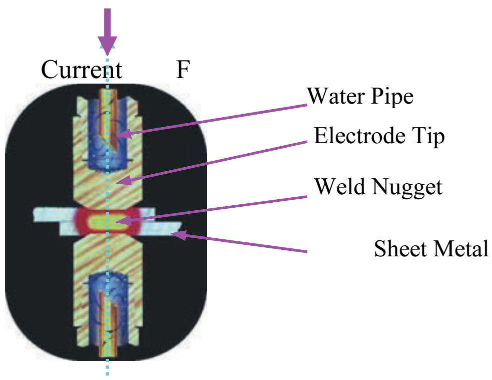


Figure 1: The spot resistance welding process

Figure 2: Tip wear samples (MPG266 GCAP)

However, bad welds are often produced due to the erosion of the welding electrode. Firstly, with usage, electrodes will "mushroom" (tip becomes blunt and contact area increases as shown in Figure 2). As the contact area increases a lower current density develops, which results in a smaller weld nugget. Secondly, when welding galvanized steel, the zirconium, chromium and zinc coating vaporize as the temperature rises, resulting in the electrode losing its hardness very rapidly and brass being formed on the electrode-weld face. This phenomenon makes the contact area sticky and larger. When an acceptable tip size and shape to form a good weld is no longer present, the electrodes need to be replaced or dressed back to the original shape.

In most modern welding production lines, there are often many weld robots varying from several to hundreds of robots. Each robot may consist of one or several stochastically failing parts with increasing failure rate, such as electrode tip in the servo gun. Generally, a weld gun on an automotive assembly line may make as high as 160,000 welds in one week [1]. In such a high productivity line, it is not cost effective to stop the whole line or change the schedule intentionally just to replace one tip. However, since robotic welding electrode tip failure time is different when performing different tasks, and the tip is relatively inexpensive compared to the cost penalty for stopping the production line, a lower cost can be achieved to replace one tip in conjunction with other tips which fall in the failure control limit than to perform them individually. In addition, most components in a robotic welding cell, i.e. servo gun and control unit etc., are subject to deterioration with usage and age, and they need to be maintained or repaired. Maintenance on them can also create a chance for replacement of other tips due for replacement. Thus, in this case, it is possible and much more cost effective to employ an opportunistic replacement policy. However, there still is a trade-off between replacing some tips together at the same time to reduce the number of line stoppages and using each tip as long as possible.

In this paper, the spot welding robot servo gun tip opportunistic replacement policy will be studied. Van-Rob, an automotive parts manufacture, has been chosen to formulate an opportunistic replacement policy, and the policy is evaluated using simulation.

2 OPPORTUNISTIC REPLACEMENT POLICY AND MODEL DESIGN

2.1 Opportunistic Replacement Policy Review

In the past 4 decades, opportunistic replacement policy has been of great interest for many real applications. The concept of Opportunistic Replacement (OR) stems from the fact that there is the possibility of dependence among various components in a multi-component system [2]. Opportunistic replacement policy means that aging constituents of a system are replaced before their normal terms while the system is down because another constituent has failed or is due for replacement. The main concern is reduced overall cost. This is particularly true in the case of a continuous system, where the failure of any one component results in stoppage of the whole system, so it is possible to do opportunistic replacement policy to non-failed subsystems at a reduced additional cost while failed subsystems are being replaced. An optimum balance between the savings obtainable from this simultaneous action and the cost of premature replacement of units that could otherwise be used for a longer period is required. Many investigators have been involved in the research of opportunistic maintenance policy.

Berg [3] suggested a preventive replacement policy for a machine with two identical components which are subject to exponential failure. Under this policy, upon a component failure the other component as well as the failed one is also replaced if its age exceeds a pre-determined control limit L. Later, Berg extended it to such a policy: both units are replaced either when one of them fails and the age of the other unit exceeds the critical control limit L, or when any of them reaches a predetermined critical age.

Zheng and Fard [4] examined an opportunistic maintenance policy based on failure rate tolerance for a system with k different types of units. Using this policy, action limits are employed in a system with repairable IFR (Increasing Failure Rate) units. When a unit is replaced, either when the hazard rate reaches L or at failure with the failure rate in a predetermined interval (L- u; L), all of the operating units with their hazard rate falling in the interval (L- u; L) are replaced at that time.

L'Ecuyer and Haurie [5] reported a dynamic programming model for the sub-optimal policy in which all failed components are replaced at scheduled inspections and also all other components whose ages exceed threshold levels at these inspections. Dagpunar [6] introduced a general policy where replacement of a component within a system is available at an opportunity. An opportunity arises if the failure of some other parts of the system allows the component in question to be replaced. Radner and Jorgenson [7] proposed an (n_i;N) policy for a series system consisting of one IFR component and M CFR (Constant Failure Rate) components. Yuna and Ferreirab [8] developed a simulation model to assess the inventory requirements of alternative rail sleeper replacement strategies. The main aim of the model is to determine the optimal replacement strategy. Pham and Wang [9] investigated the opportunistic maintenance of a k-out-of-n system with imperfect preventive maintenance (PM) in aircraft engine maintenance.

From the above review, it is clear that most of the research work on opportunistic maintenance is concentrated around (n;N) type policies. Nearly all of these consider single OM age for each of the IFR components irrespective of which other component is taken down. When the number of IFR components increase, just like robotic welding line which consists of so many robots and failure time for different components is different, the complexity of the model increases. No successful application in robotic welding production has been reported.

2.2 Replacement Policies Design for Robotic Welding Electrode

Each replacement policy has its specific advantages and disadvantages. The best replacement policy depends on the particular circumstance. The policy considered here is an (n_{ij};N_{ij}) policy based on tip failed and classification of opportunities.

Where

j, the opportunity class

n_{ij}, the OR age for tip i when the system is taken down for replacing tip j

N_{ij}, the OR age for tip i at which PR or CR is performed on component j

Therefore, three types of replacement policies have been proposed for robotic servo gun maintenance and some of them have already been employed.

Corrective Replacement Policy (CRP)

The tip is replaced immediately by a new one upon sudden failure. The phenomenon does happen sometimes when the current is too high. Corrective replacement is still one of the most commonly used replacement methods in industry. This policy should be avoided if possible due to its expense. It is a random variable but it does create opportunity for other tips' replacement.

Preventive Replacement (Maintenance) Policy (PRP)

Components are maintained periodically to avoid abrupt breakdowns, which carry great penalty. Research indicates [10] over 80% of all weld guns are repairable up to three times before a new gun purchase is necessary. Servo gun repair or replacement can include all parts showing wear or damage: such as air cylinders, transformer, gun arms, water cooled cables, etc. In this case, the replacement or maintenance interval is a decision variable which is based on the life expectancy of each component and production process economy. For robotic spot welding, the maintenance records based on number of welds made will be a good guide. One of the real experience suggested that every 125,000 welds could provide an opportunity to check and replace tips as needed [1].

Opportunistic Replacement Policy (ORP)

Replace all the failed tips (active replacement) either when the age reaches T or its next failure time falls in a predetermined interval. Namely, for the same predetermined interval L of all tips, when a tip is replaced due to the tip wear reaching T, all of the operating tips in this production line with their tip wear falling in the interval (T; T+L) are replaced at the same time.

They can also be replaced in groups at repair or replacement of other components, each shift change, Maintenance or retooling of the production line as shown in Table 1. In this project, the electrodes are monitored continuously, so the next expected failure time of each tip can be easily obtained. Then the OR ages for each tip can be obtained corresponding to each opportunity class, irrespective of the component taken down for CR or PR or OR. In this study, minimization of times of line downtime, which consists of the main part of total cost, has been considered as a main objective.

Table 1

Work schedule for production line

No	Time	Total (min)	Description
1	7:00 – 7:10	10	Idle
2	7:15 – 7:24	11	Maintenance on GEO 2
3	9:25 – 9:35	10	Tip change
4	9:37 – 9:41	4	Maintenance
5	10:18 – 10:38	20	Maintenance on GEO 1
6	11:35 – 12:00	25	Break
7	12:00 – 12:25	25	Unscheduled
8	12:43 – 12:53	10	Tip change
9	13:50 – 14:00	10	Break
10	14:00 – 14:10	10	5S
11	14:35 – 14:50	15	Maintenance on GEO 1
12	15:20 – 15:28	8	Tip Change
Total	8.5 hour	158	

Notes:5S(Separate, Simplify, Sanitize Standardize and Support

3 EXPERIMENTAL SETUP AND DATA ANALYSIS

3.1 Industrial Partner Van Rob.

The experimental development and implementation phase of this work has been conducted on the robotic resistance spot welding system, which consists of 6 Nachi SH166-01 robots or GEO Station robots which are shown as Figure 3, at Van Rob. The product concerned in this experiment is an automotive part consisting of 4 sequences [11].

Figure 3: Van-Rob robotic welding system

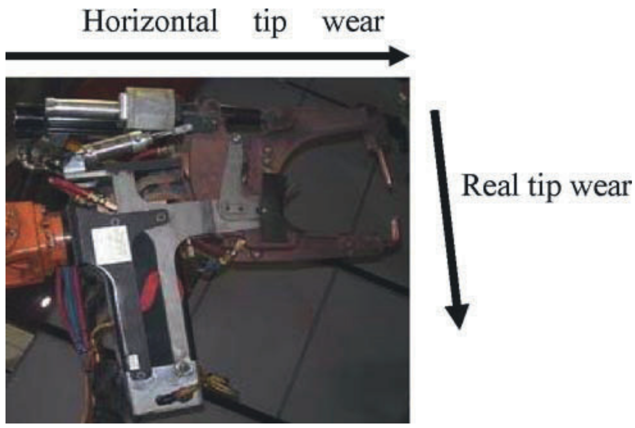


Figure 4: Servo gun tip wear compensation methodology

3.2 Tip War Monitoring and Compensation Methodology

As welding continues, the soft copper mushrooms resulting in a larger distance between the upper and lower tips. This should be compensated for, to guarantee a good weld. Therefore, Van-Rob has employed a monitoring system to realize the tip wear control. That is, the welding robot performs a tip check with no current passing through after welding several sequences (welds). The process is automatically compensated by a preset horizontal gun holder forward value (4 at the beginning in this case) shown in Figure 4. This value is automatically collected and input into the PLC, and then transferred to an internal database. Although it needs to be transferred to the real tip wear, it still could be a good indicator for current application when the real tip wear is not measured directly.

3.3 Tip Wear Data Analysis

Tip wear model and band limits

The real field data have been obtained on the servo gun of robot LHsub1. 21 tips have been used in this period. Data sample for tip 18 is shown in Table 2. The relationship of the tip wear compensation value with welding number is summarized and shown in Table 3.

The results show that the total weld number is around 1450 welds and total tip wear compensation amount will be about 2.1-2.6mm for each tip. The data also indicates that there exists a similar tip wear trend. It is therefore, possible to find a suitable model for all tips. The behaviors of the refined curve fitting models show that 4th and 7th order polynomial functions work very well for nearly all tip shifts. After averaging the models of the representative tip shifts, the general model of tip wear can be obtained and expressed as equation 1. Where x equals to welding count divided by 100. The tip wear trends, its general model and band limits behaviors are shown in Figure 5.

$$Y(x) = -5.4431e-005x^7 + 0.0029115x^6 - 0.062392x^5 + 0.69198x^4 - 4.3739x^3 + 16.984x^2 - 54.363x + 520.16 \quad (1)$$

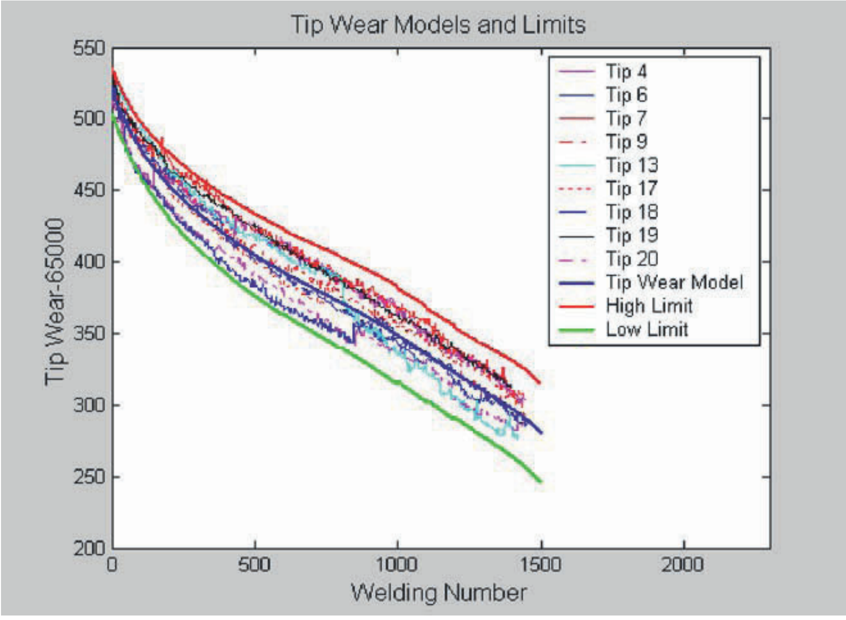


Figure 5: Tip wear model with its band limits

Table 2

Data sample for tip 18

Date &Time	Log #	Part #	Sequence #	Tip Wear(*100)	Note
2003-6-12 4:54	4761	210	1	65519	
2003-6-12 4:54	4762	210	2	65519	
2003-6-12 4:54	4763	210	3	65519	
2003-6-12 4:54	4764	210	4	65519	
2003-6-12 4:54	4765	210	24	65519	Tip Check
2003-6-12 4:55	4766	211	1	65515	
2003-6-12 4:55	4767	211	2	65515	
2003-6-12 4:55	4768	211	3	65515	
2003-6-12 4:55	4769	211	4	65515	
2003-6-12 4:55	4770	211	24	65515	Tip Check

Table 3

Relationship of the Tip Wear versus welding number

Tip No.	Total weld number	Tip Wear value (Staring from)	Tip Wear value (Ending at)	Total Tip Wear (compensated)
4	1448	65525	65304	221
6	1464	65520	65287	233
7	1441	65533	65306	227
9	1426	65530	65304	226
13	1424	65535	65276	259
17	1452	65531	65294	237
18	1460	65519	65278	241
19	1401	65531	65314	217
20	1452	65513	65282	231

3.4 Tip Time and Weld Order

The life time versus weld count for weld tip sample are shown as Figure 6 and listed in Table 4.

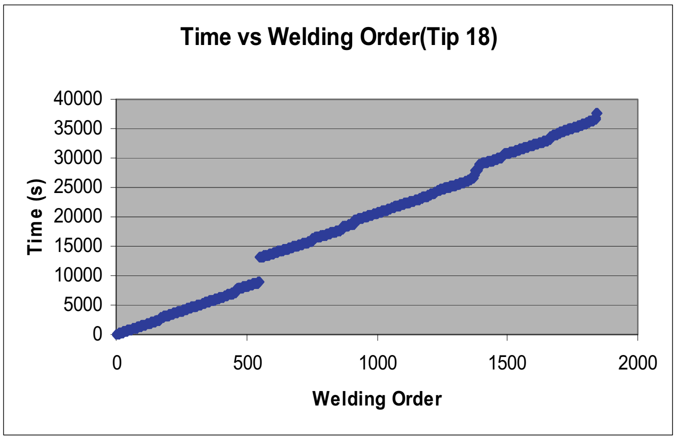


Figure 6: Time vs welding order

Table 4

Total time and time per weld

Tip No.	Total time (hours)	Time per weld (sec)
18	10.48	20.46
2	9.064	17.98
12	10.82	21.26
4	9.85	19.08
Average	10.05	19.69

From the above results, most tips last about 9 to 11 hours and each weld takes about 19.696 seconds for this gun. The real weld time should be considerably less due to the fact that there is significant idle time between welds and sometime is used to move the gun or rotate the parts to get to the next weld. It is also clear that the life expectance of each tip even in the same gun is different, but generally they may have a distribution (i.e. Weibull Distribution). Based on the average weld time, real tip wear and band limits, we can obtain the recommended and maximum number of welds left and the corresponding remaining lifetime for a specific tip.

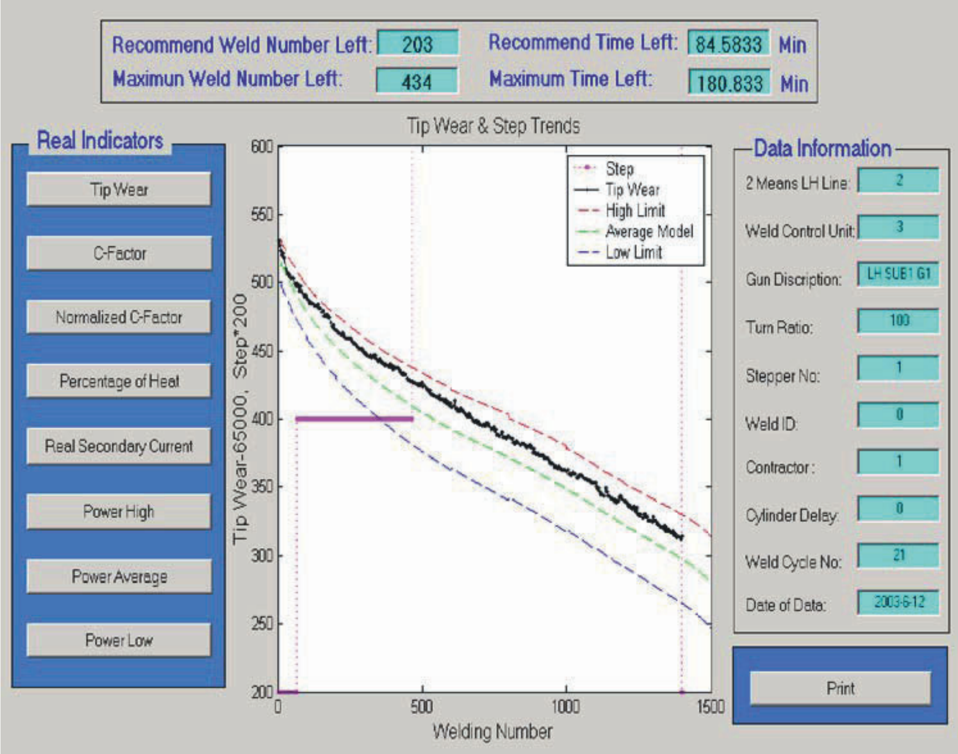


Figure 7: Tip wear next failure time prediction interface

4 OPPORTUNISTIC REPLACEMENT POLICY SIMULATION

Although many mathematical models exist, Tip Opportunistic Replacement cannot be easily modeled using mathematical models. In nearly all the case studies of real-life situations, the simulation approach has been adopted. It has been demonstrated to perform well in tackling maintenance problems and it can also model the multi-component system with relative ease. Therefore, the simulation approach is adopted in this paper.

4.1 Assumptions

For simulation simplicity, the following assumptions have been made:

1). Only the electrode tip is an Increasing Failure Rate (IFR) for each robot. Other components are constant failure rate (CFR) and no corrective replacement. The tip wear status is known and at no cost due to the condition monitoring.

2). Every tip failure causes production line failure and needs replacement right away.

3). The time to failure of each new tip under normal working condition follows the similar tip wear trend obtained and simulation starts at a fixed failure time for each robot.

4). A replacement constitutes the installation of a new tip or the removal of the current tip. However, replacement time is negligible compared to times between failures.

5). Cost assumption. All replacement activities are performed by permanent employees. Manpower cost is fixed. For this continuously operating system, downtime costs are more predominant. For simplicity, each CR, OR and PR downtime for specific components are assumed to be constant although they will vary over a wide range depending on the amount of work done. Among these three costs, CR cost is the highest because it will carry a cost penalty when a component suddenly fails. PR and OR costs are the same if tip condition monitoring is performed well. Based on the production economy, production line maintenance parameters are listed as Table 5.

For each specific replacement interval control limit L, the total cost of the replacement for adopting OR is as follows.

Table 5

Production line maintenance parameters

Parameters	Symbol	Value
Tip Life distribution	T	$W(\alpha, \beta)$
Corrective replacement cost	C_{cr}	$10/each
Preventive replacement cost	C_{pr}	$5/each
Opportunistic replacement cost	C_{or}	$5/each
Loss of production	C_{loss}	$25/each
Cost of electrode	C_{tip}	$1.2/each
Cost of tip working productivity	C_{waste}	$20/hour

$$TC_{new}=N_{cr}*C_{cr} + N_{or}*C_{or} +N_{pr}*C_{pr} +N_{tip}*C_{tip}+C_{loss}*(N_{cr}+N_{or}+N_{pr}) \quad (2)$$

The total cost of the replacement if not adopting OR is as follows (Each time need to stop line for tip replacement or preventive maintenance).

$$TC_{old} = N_{tip}*(C_{tip}+C_{loss})+(N_{cr}+N_{pr}) *C_{loss} \quad (3)$$

Then the saving cost between adopting OR method and not adopting OR method will be

$$TC_{saving} = TC_{old} -TC_{new} \quad (4)$$

The total money wasted, due to the tip life wasted that could be used for more welds because of adopting OR method, will be

$$TC_{waste} =Sumtime_{waste}*C_{waste} \quad (5)$$

Therefore, the final saving between money saved and wasted after adopting the OR will be

$$TC_{finalsave}=TC_{saving} -TC_{waste} \quad (6)$$

Where:

N_{or} : Number of Opportunistic Replacement times

C_{or} : Each Opportunistic Replacement cost

N_{cr}: Number of Corrective Replacement times

C_{cr} : Each Corrective Replacement cost

N_{pr}: Number of Preventive replacement times

C_{pr} : Each Preventive Replacement cost

N_{tip}: Total number of tip replacement

C_{tip}: Cost of each tip

C_{loss}: Cost penalty for loss of production

TC_{old}: Total cost of the replacement if not adopting OR

TC_{new}: Total cost of the replacement if not adopting OR

TC_{saving}; Saving cost between adopting OR and not adopting OR

TC_{waste}: Total money wasted due to still useable tip life wasted

$Sumtime_{waste}$: The total useable tip life wasted that could be used for more welds

C_{waste}: Cost of tip working productivity (for lose of tip life due to OR)

$TC_{finalsave}$: Final saving between money saved and wasted after adopting the OR

From the final saving for each control limit, we can obtain the optimum replacement interval by finding the maximum savings. Namely:

$$\frac{\partial TC_{finalsave}}{\partial L} = 0 \tag{7}$$

4.2 Simulation and Results

The simulation interface developed is shown as in Figure 8. Each expected robotic tip life can be changed directly or by the controller according to the real situation. The simulation system can be simulated for any period of time with the imposition of the policy after selecting the simulation time length.

Results for any interval control limit

First, for any gradually increased replacement interval L, this simulation process is executed within one year. Irrespective of the component taken down for CR or PR or OR, all the operating tips whose age reaches its life expectance T or the difference between next replacement time and current age is less than the replacement interval L , have been replaced at the same time. And then the tip replacement and cost are obtained. Using this simulation approach, it is easy to get the cost trends and obtain the optimum replacement interval to save the maximum amount of money. After simulation with opportunistic replacement policy, the total number of replaced tips and line stoppage, total saved cost, total wasted cost due to the wasted tip life and final money saved are obtained and shown in Table 6 for each specific control limit. Each robot's next replacement age is also displayed on the interface.

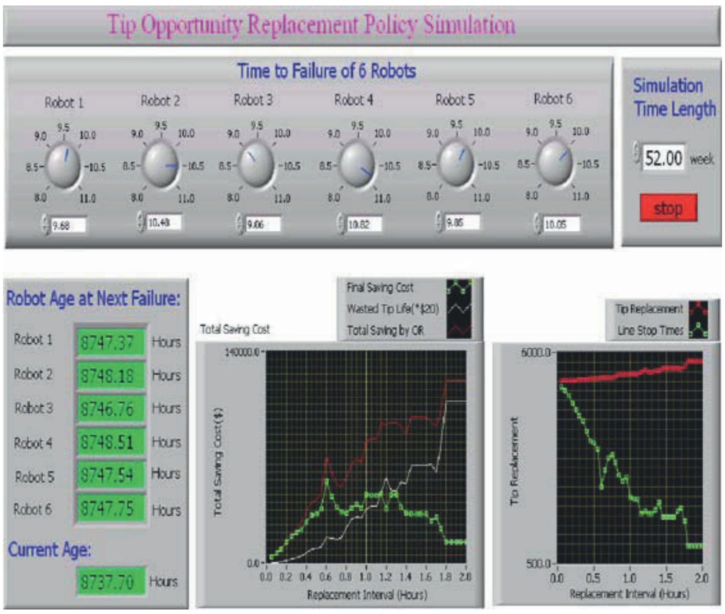


Figure 8: Simulation result interface

From the table, it is clear that the replacement interval L plays a significant role in tip opportunistic replacement due to the fact that it decides how many tips can be replaced at a time and how many times it needs to stop the production line. Both of these contribute a lot to the total cost. As the replacement interval L becomes larger, the total saved cost, total wasted cost and the total number of replaced tips all increase, but the total number of line stoppages for OR replacement deceases significantly. All of these trends become constant when the interval limit reaches 1.8 hours, (namely replacing all tips each time). Based on the cost assumption, the total money saved is obtained for each specific control limit and the results show that it will increase first and then decrease. There does exist a maximum value at interval control limit of 0.6 hours. Up to $53,944 per year can be saved after employing an opportunistic tip replacement policy for this 6 robot system. Currently there are 72 Servo guns working continuously at Van Rob, some of them need to replace tips at around 2.5-3 hours or even less than 1 hour. It is also estimated that only the tip purchase cost at Van-Rob is $500,000/year. Therefore, it is expected to save more than $300,000 /year after employing the opportunistic replacement policy for all robot cells.

Results under constant interval limit but with increasing welding time

The trend under constant interval limit but with increasing welding time is also simulated. The optimizing interval limit (L=0.6hr) is chosen and the results are shown in Figure 9 and Table 7. From the above figures, it is clear that the total expected cost, the total stop line times and the total tips being replaced are increased linearly with the welding time. Therefore, it will save money in the long term

Table 6

One year simulation results for 6-robot welding system

Simulation interval L(hr)	Total tips been replaced(TNR)	Total stop line times (TST)	Total expected cost save using OR($)	Note
0.10	5263	5004	6204	
0.20	5271	4584	15428	
0.30	5278	4370	19684	
0.40	5299	3827	29,569	
0.5	5309	3619	32,698	
0.6	5350	2502	53,944	Maximum saving
0.7	5348	2638	50,684	
0.80	5398	2827	37,063	
0.90	5417	2703	37,000	
1	5440	2227	45,065	
1.10	5451	2177	44,325	
1.20	5500	1780	45,173	
1.30	5500	1780	45,173	
1.40	5523	2106	32,990	
1.5	5585	1713	32,443	
1.6	5585	1713	32,443	
1.7	5652	1615	22,707	
1.80	5790	965	12,931	
1.90	5790	965	12,931	
2	5790	965	12,931	

Table 7

Simulation results of tip OR under different welding time (Interval Limit L=0.6 hour)

Welding time (Week)	Total number of tips replaced	Total stop line times	Total expected final saving ($)
1	102	49	1,006.2
4	410	193	4,112.6
26	2673	1251	26,935
52	5350	2502	53,944

Note: Based on 168hours /week.

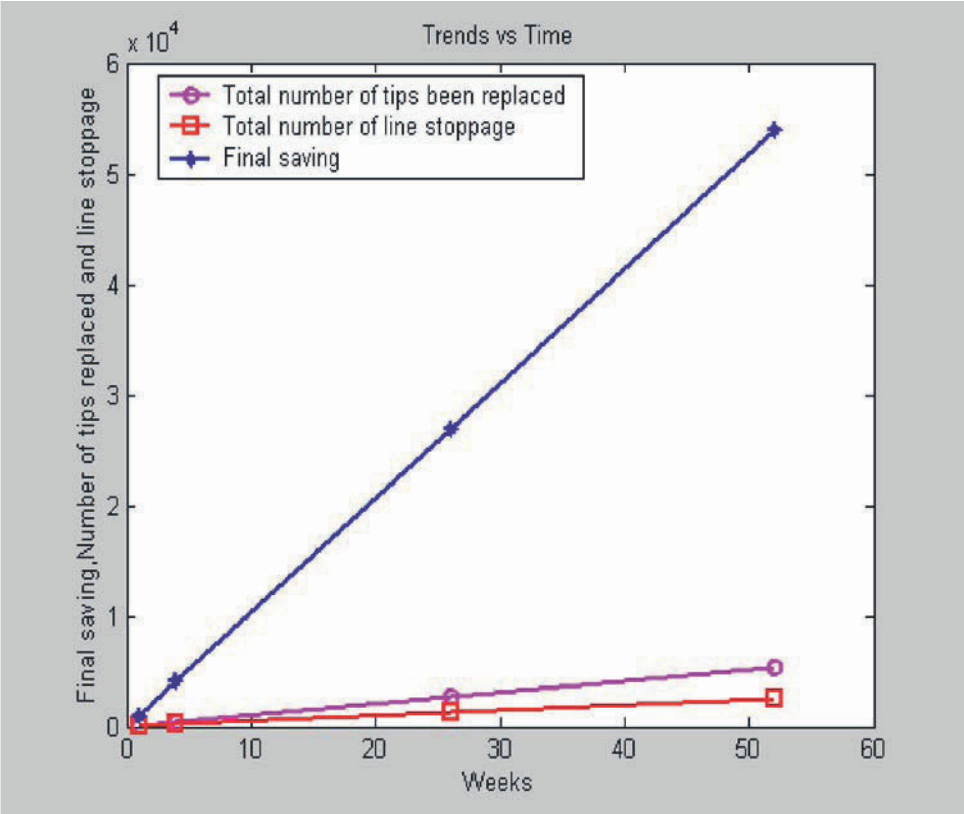


Figure 9: The result trends under L=0.6hour

5 SUMMARY

This paper addresses the electrode replacement problem for a robotic resistance spot welding system consisting of 6 non-identical robots and the general maintenance optimization problem in a manufacturing setting. The electrode tip wear compensation methodology for tips that mushroom as the number of welds increases is introduced. Three kinds of replacement policies are described. They include; corrective replacement, preventive replacement and opportunistic replacement. Polynomial models have been used to predict future tip behavior by modeling the tip wear from field data. The results show that a 7th order polynomial model can represent the tip wear trend. Welding order and time analysis results show that most tips last about 9-11 hours and each weld takes near 20 seconds. Based on this, a tip wear trend as well as high and low limits for tip usage were established. Subsequently, the recommended and maximum number of welds left on a given tip and the corresponding remaining lifetime for a specific tip can be obtained.

The expected failure time or replacement time can also be predicted. Through a careful cost investigation, a simulation model for electrode tip opportunistic replacement on a 6-robot resistance welding system has been developed. Based on the assumptions, the first simulation results show that there do exist a maximum saving value at interval control limit of 0.6 hours. Up to $53,944 per year can be saved after employing for tip opportunistic replacement for this 6 robot system under such a situation. The second simulation shows that under increasing welding time, the total expected cost, the total stop line times and the total tips been replaced are increased linearly. It is concluded that it is worthwhile to consider the an Opportunistic Replacement Policy.

6 ACKNOWLEDGEMENTS

Thanks are extended to Van-Rob for providing the great support on this project. We further thank Dr. Kornel Farkas and Mr. Jeffrey Liaw for providing their help in the experimental work. We also appreciate Mr. Brian Morris and Dr. Sam Li for their valuable information and suggestions.

7 REFERENCES

[1] http://www.updatetechnology.com

[2] Neelakanteswara A and Bhadury RB, (200) Opportunistic Maintenance of Multi-equipment System: a Case Study. *Reliability Engineering International*, Vol.16, Issue 6, pp.487-500.

[3] Berg M. (1976) Optimal Replacement Policies for Two-unit Machines with Increasing Running Costs. *Stochastic Processes and Applications*, Vol. 5, pp.89–106.

[4] Zheng X and Fard N. (1992) Hazard-rate Tolerance Method for an Opportunistic Replacement Policy. *IEEE Transactions on Reliability*, Vol. 41, Issue. 1, pp.13-20.

[5] L'Ecuyer L and Haurie A. (1983) Preventive Replacement for Multicomponent Systems: an Opportunities Discrete Time Dynamic Programming Model. *IEEE Transactions on Reliability*, Vol. 32, Issue. 1, pp.17 –18.

[6] Dagpunar JS. (1996) Maintenance Model with Opportunities and Interrupt Replacement Options. *Journal of the Operational Research Society*, Vol. 7, pp.1406–1409.

[7] Rander R and Jorgenson DW. (1963) Opportunistic Replacement of a Single Part in the Presence of Several Monitored Parts. *Management Science*, Vol. 10, pp.70–83.

[8] Yun WY and Ferreira L. (2003) Prediction of the Demand of the Railway Sleepers: A Simulation Model for Replacement Strategies. *Int. J. Production Economics*, Vol. 81, pp.589–595.

[9] Pham H and Wang H. (2000) Optimal (τ; T) Opportunistic Maintenance of a k-out-of-n: G System with Imperfect PM and Partial Failure. *Naval Research Logistics*, Vol. 47, pp.223–239.

[10] Profile, Stillwater Technologies Inc. 2003

[11] Delta line investigation Profile, Van-Rob, 2003, http://www.van-rob.com

[12] Mohamed-Salah O. (1999) A Simulation Model for Opportunistic Maintenance Strategies, *Proceedings of the Annual Reliability and Maintainability Symposium*, IEEE, pp. 703-708.

[13] Zheng X. (1995) All Opportunity-triggered Replacement Policy for Multiple-unit Systems. *IEEE Transactions on Reliability*, Vol. 44, Issue. 4, pp.648 –652.

[14] Degbotse AT and Nachlas JA. (2003) Use of Nested Renewals to Model Availability under Opportunistic Maintenance Policies. *Proceedings of the Annual Reliability and Maintainability Symposium*, IEEE, pp. 344 - 350.

INDUCTION MOTOR FAULT DETECTION USING HYBRID METHODS

Li W and Mechefske CK

Department of Mechanical and Materials Engineering, Queen's University, Kingston, Ontario, Canada K7L 3N6

Abstract: This paper presents the comparison results of induction motor fault detection using stator current, vibration and acoustic methods. A broken rotor bar fault and a combination of bearing faults (inner race, outer race and rolling element faults) were induced into variable speed three-phase induction motors. Both healthy and faulty signatures were acquired under different speed and load conditions. Spectral differences are compared and characterized under constant speed conditions. Experimental results show that the stator current method is sensitive to the broken rotor bar fault while the vibration method is sensitive to bearing faults. The acoustic method is very attractive in that it contains less noise and interference within the analyzing frequency band. With the proper selection of monitoring and analysis methods, induction motor faults can be detected accurately under both stationary and non-stationary states.

Key Words: Induction Motors, Motor Current Signal Analysis, Vibration Signal Analysis, Acoustic Signal Analysis, Varying Speed, Varying Load.

1 INTRODUCTION

Over the past few decades, condition monitoring techniques of induction motors have been extensively conducted with an emphasis on motor current signature analysis (MCSA). MCSA, a noninvasive monitoring technique, is sensitive to most motor faults. As a result, a large amount of research has been directed toward the analysis of motor current [1-3]. Besides MCSA, vibration analysis is also an effective monitoring technique in detecting motor bearing and broken rotor bar faults [4-5].

The signal analysis methods range from time domain statistical analysis to joint time-frequency analysis and include the uses of artificial neural network and wavelet transforms [6]. Spectral analysis of motor current signatures was performed by Benbouzid et al. [7]. Several motor asymmetrical faults were compared in the frequency domain. Higher-order spectral analysis was also applied to detect motor faults [8]. The analysis is effective in detecting phase couplings which occurs when motors have faults. Yaz c and Kliman [9] used the short time Fourier transform (STFT) as a pre-processing method to extract fault features from motor current signatures. Filippetti et al. reviewed the applications of artificial intelligence (AI) methods in the diagnosis of motor faults and analyzed the factors influencing the current components [10]. The analysis provided insights in calculating the amplitudes of current, torque, and speed spectral components in a transient model. Ye at al. [11] applied a wavelet packet decomposition (WPD) method to analyze the characteristics of the stator current under normal and faulty conditions. The feature coefficients extracted from the WPD were then used detect rotor bar breakage and air-gap eccentricity faults.

By reviewing the past work it is noted that although advanced signal processing methods have been extensively employed, conventional statistical calculations and Fourier series based spectral analysis still provide vital preliminary information. For instance, both the neural network and model based methods rely heavily on the calculations of motor signatures (e.g. RMS, Kurtosis, etc.) and fault frequencies (via FFT). Therefore, understanding about the characteristics of motor related faults under different speed and load conditions is desirable in order to extract efficient features.

To address the above issue, this paper is intended to make comparisons of motor fault detection capabilities using different methods. In addition to the widely used motor current and vibration methods, an acoustic method, a noninvasive monitoring scheme, is also employed in this paper. Comparisons are to be made under different load and speed conditions in the frequency domain.

2 MOTOR FAULTS

As described in the previous section, failures of motor core components such as stators, rotors and bearings account for a large percentage of motor breakdowns. Hence it makes more sense to simulate the most common motor faults in an experimental setting. In this paper, a broken rotor bar fault and a combination of rolling element bearing faults are induced experimentally.

2.1 Broken Rotor Bars

Rotor failures account for 5-10% of induction motor problems. A broken rotor bar fault has distinctive characteristic frequencies which can be calculated as

$$f_{brb} = (1 \pm 2ks) f_0 \quad , \; k = 1,2,3,\cdots \tag{1}$$

Where f_0 is the driving frequency and s is the fractional slip of the motor.

Once there is a broken rotor bar fault, sidebands around the driving frequency can be expected in the power spectrum. In particular, the first-order sidebands (e.g. $k = 1$) are of particular interest in the detection of broken rotor bar fault. The left sideband $(1-2s) f_0$ is due to electrical or magnetic rotor asymmetry caused by broken rotor bars while the right sideband $(1+2s) f_0$ is due to the speed ripple or variation [10]. The amplitudes and presence of the sidebands depend on the physical position of the broken rotor bars, speed and load. The locations of the sidebands will shift outwards as the speed and load are increased. It is acknowledged that the sidebands may also be observed when the motor has no broken rotor bar fault as rotor ellipticity and shaft misalignment could both induce rotor asymmetry to a certain extent [9]. However, the sideband amplitudes produced in those cases are much smaller compared with those produced by broken rotor bar fault. Benbouzid suggested that when the amplitudes of these harmonics are over 50 dB smaller than that of the fundamental frequency component, the motor could be considered healthy [3]. Such a judgment needs to take into account the speed and load factors. In this paper, a motor with one broken rotor bar was pre-manufactured and used in the tests to simulate the broken rotor bar fault.

2.2 Bearing faults

Of the large power machine failures, bearing problems account for over 40% of machine breakdowns. Failure in rolling element bearings is often a gradual process starting from fatigue fissures located beneath the bearing raceways. As a result of the continued concentration of stress, the fissures propagate toward the surfaces resulting in minute cracks. These cracks, under cyclical loading, develop into large cracks and cause the surface material to break loose in the form of flaking and spalling. Other causes of bearing failure include improper installation, lubrication, contamination and corrosion.

Rolling element bearing fault frequencies include ball pass frequency outer race (BPFO), ball pass frequency inner race (BPFI), ball spin frequency (BSF), and fundamental train frequency (FTF). All of these frequencies can be calculated from bearing geometric dimensions. In our tests, a combination of bearing faults including inner race, outer race and rolling element faults were seeded in a SKF bearing 6203. Table 1 lists bearing geometric dimensions and fault frequencies.

Table 1

The test bearing specifications and frequencies.

Bearing model	SKF 6203
Number of balls	8
Bore diameter	17 mm
Outside diameter	40 mm
Ball diameter	6.75 mm
BPFI	4.95
BPFO	3.05
BSF	1.99 or 3.99
FTF	0.382

The bearing frequencies listed in Table 1 are related to inner shaft rotation frequency in Hertz. For instance, if the shaft is rotating at 1 Hz (60 RPM), the BPFO is 3.05 Hz (3.05×1 Hz). Two BSFs are given in Table 1. The first (1.99) is called the rotation frequency of rolling element and the other is called the over-rolling frequency of one point on a rolling element which occurs when an element has defects. However, it should be noted that, unlike BPFO or BPFI, BSF may not always be present as the ball may change its rotational axis and thus not always be over-rolled at the same point.

3 EXPERIMENTAL METHODS

Experimental tests were conducted in a laboratory environment using a dynamic fault simulator. The system consists of an induction motor, variable speed controller, a gearbox and a magnetic brake unit. Figure 1 illustrates the simulator system and the associated data acquisition instrumentation.

The test motor is a two-pole three-phase induction motor with rated output power of 3 HP (2.24 kW). To change the speed of the motor, a digital variable speed controller is attached to the test rig between the power line source and the AC motor. The controller can be programmed to any specific shaft rotation speed between 0 and 3600 RPM. A gearbox is used to reduce the output shaft rotational speed. To apply the load, a magnetic brake unit is employed on the output shaft of the gearbox. The brake unit has an adjustable load range of 0-70 lb-in (0-7.9 Nm).

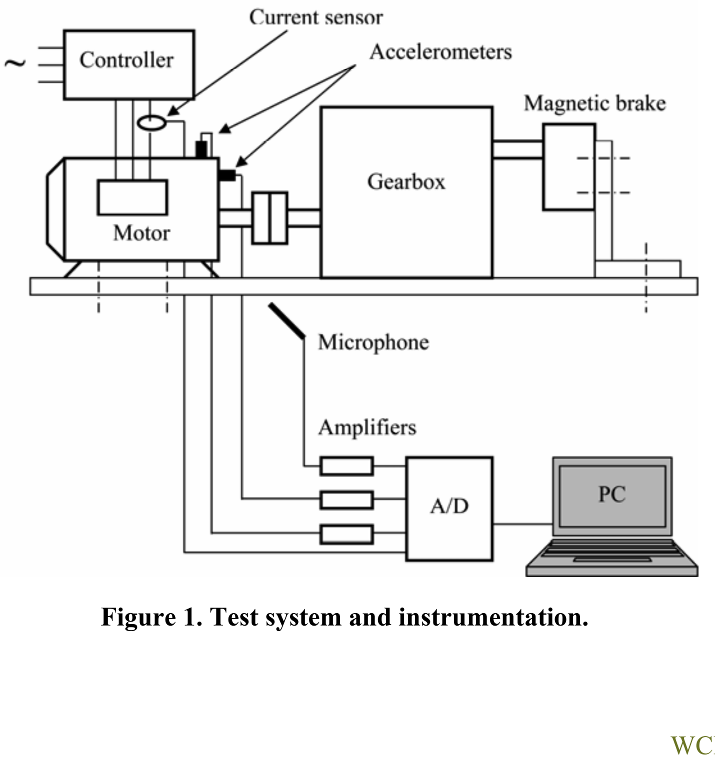


Figure 1. Test system and instrumentation.

One Hall-effect current sensor, two accelerometers and a microphone were used to acquire the stator current, vibration and acoustic signatures. One accelerometer was placed on the top of the motor in the radial direction and the other was glued to the right-hand side surface along the axial direction. The microphone was pointed to the center of motor shaft and positioned about 15 cm away. Three motor conditions were simulated in the tests: normal, one broken rotor bar and a combination of bearing faults on the inner race, outer race and a ball. The tests were conducted under 5 speed and 4 load conditions. The speeds were increased from 15 Hz up to 55 Hz with a step of 10 Hz each time. At each speed, four loads were applied to the gearbox output shaft at the percentage of 0%, 33%, 66% and 100% of the maximum 70 lb-in.

4 EXPERIMENTAL RESULTS

In this section, the experimental results are presented in three subsections in the order of broken rotor bar fault, bearing failure.

4.1 Broken rotor bar

Figures 2 to 5 show the spectral comparisons of vibration, current and acoustic signatures around the driving frequency under normal and broken rotor bar fault conditions at 25 Hz (1500 RPM) and 100% load. The radial vibration spectra in Figure 2 show the sideband differences between normal and broken rotor bar on both sides of the driving frequency (25 Hz). Under the normal condition, amplitudes of the first-order sidebands are around -28 dB (left) and -30 dB (right) respectively. Under the faulty condition, the sideband amplitudes are increased to -19 dB and -21 dB, a difference of 9 dB. The axial vibration spectra in Figure 3 display the first-order left sideband difference of around 5 dB while the first-order right sidebands are almost the same. This fact indicates that the radial vibration is more sensitive to the broken rotor bar fault than the axial vibration. It is observed from Figures 2 and 3 that the frequency positions of the maximum vibration spectral components were shifted leftward to about 24.4 Hz. This is because the accelerometer signals are linked to the speed of the rotor which is slower than the speed of the magnetic fields (25 Hz in this case) due to the effect of slip.

The comparison of the current spectra in Figure 4 shows more distinctive differences. The differences in the first-order sideband amplitudes are clearer than those of the vibration spectra. The sideband amplitudes of the broken rotor bar are 9.6 dB (left) and 11.3 dB (right) higher than those of the normal condition sidebands. Compared with vibration spectra, it is noted that the sidebands under the normal condition are more clearly visible. The current spectra also display the amplitude differences at other higher-order sidebands. For instance, both the second (25 ± 2.2 Hz) and third-order (25 ± 3.2 Hz) sidebands under the broken rotor bar condition are comparatively higher than those of the normal condition sidebands.

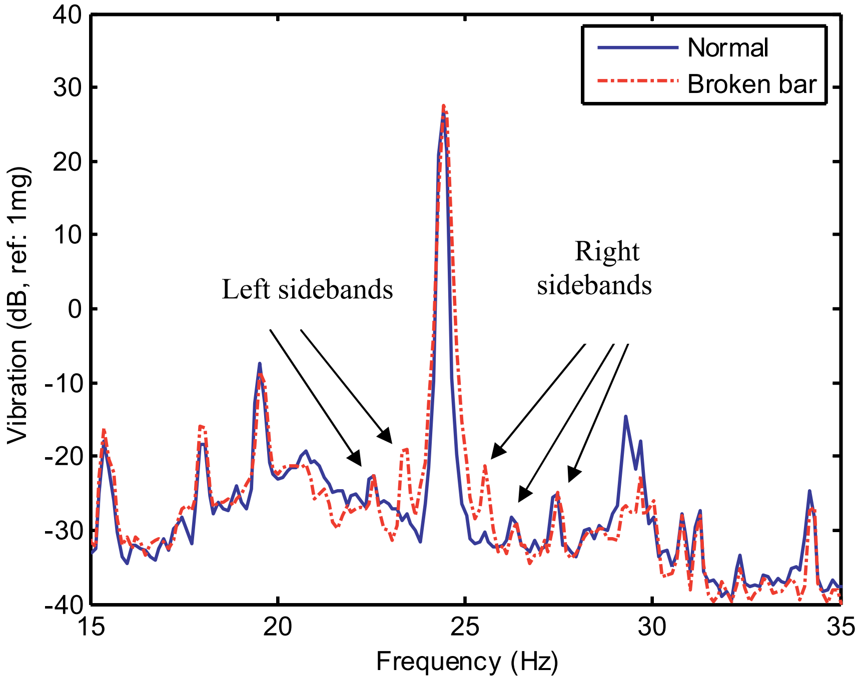


Figure 2. Radial vibration spectra of normal and faulty motors

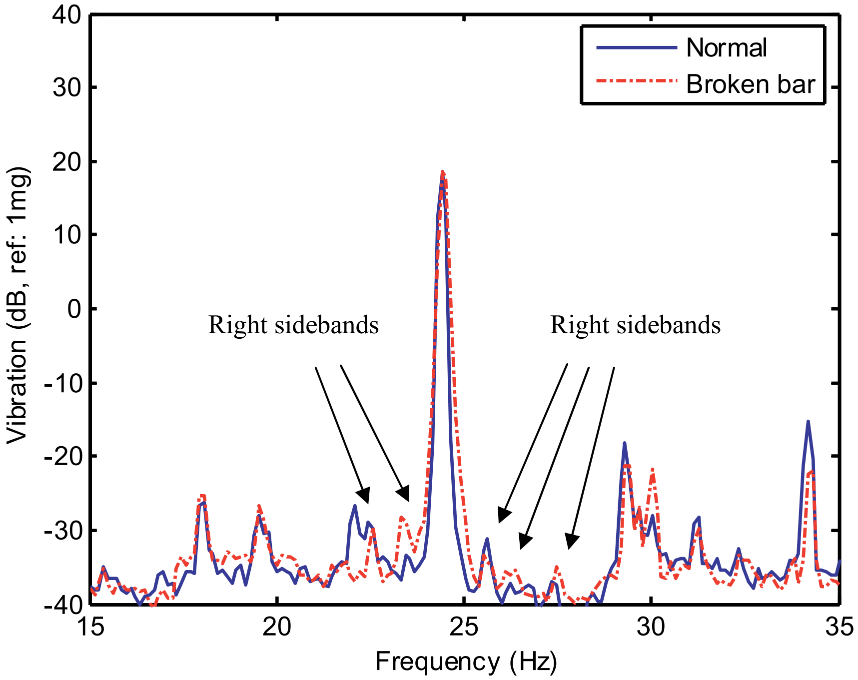


Figure 3. Axial vibration spectra of normal and faulty motors

The acoustic spectral comparison is given in Figure 5. First, the amplitude of the 25 Hz driving frequency is not as distinguished as those in vibration and current spectra. This may be explained by the fact that the sound generation mechanism is different from that of vibration. The 25 Hz driving frequency would not necessarily produce the loudest sound pressure level (SPL). Second, it is noted that the overall sound pressure level of the faulty spectrum (dash dot line) is higher than that of the normal spectrum. However, such a difference in overall spectrum amplitudes is not consistent in other types of faults. For instance, the faulty bearing acoustic spectrum is lower than the normal bearing spectrum. This is clearly because different motors have different baseline acoustic levels that affect the overall sound pressure levels. It is noted from Figure 5 that both the normal and faulty spectra show first few order sidebands. The faulty sidebands are relatively higher than the normal sidebands as a result of its higher spectral level. The differences between the first-order sidebands are 0.5 dB (left) and 1.7 dB (right).

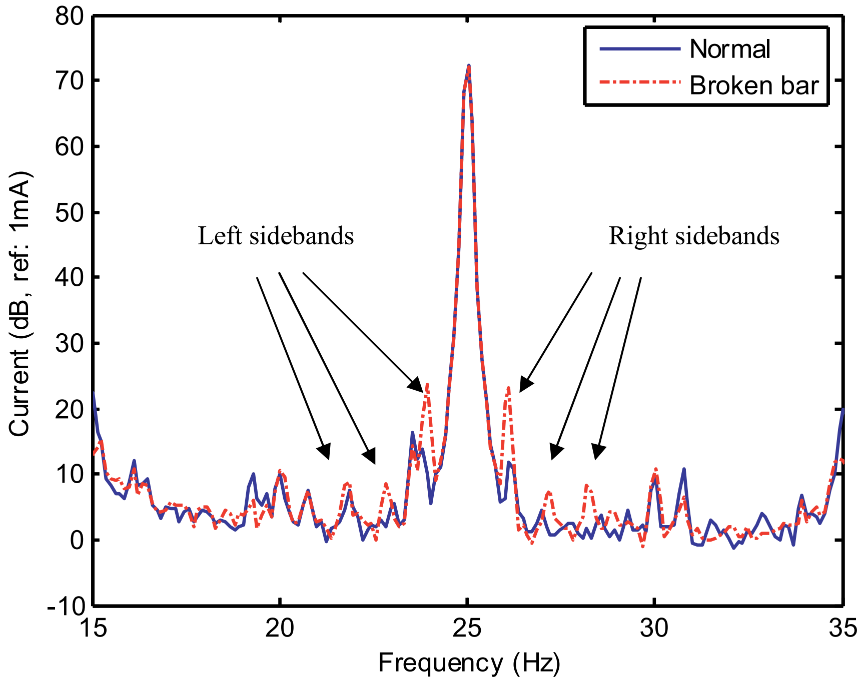


Figure 4. Current spectra of normal and faulty motors

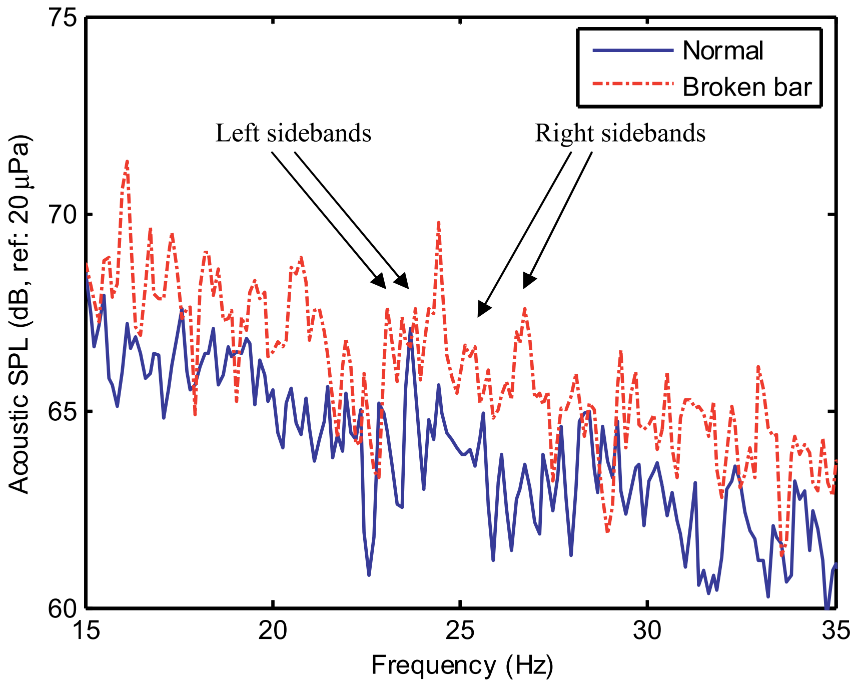


Figure 5. Acoustic spectra of normal and faulty motors

The influence of load on broken rotor bar induced sidebands is illustrated in Figure 6. Figure 6 shows the comparisons of both the left and right sideband amplitudes between normal and broken rotor bar at 55 Hz and under 4 load conditions (e.g. 0%, 33%, 66% and 100% of 70 lb-in.). First, the two vibration spectra in Figures 6 (a) and (b) show that the amplitude differences of the left sidebands are more distinctive that that of the right sidebands. Second, the current sidebands in Figure 6(c) show a gradual increase in amplitude differences against the load. Clearly, the sidebands from motor current signatures are more sensitive to load variations than those from the vibration signatures.

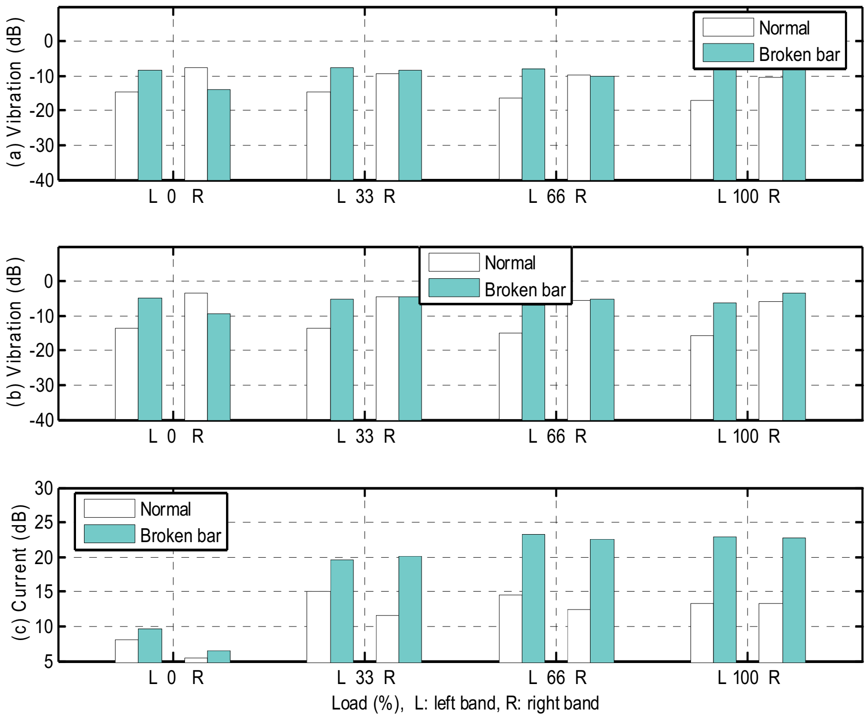


Figure 6. Influence of load on sideband amplitudes at 55 Hz; (a) radial vibration; (b) axial vibration; (c) current.

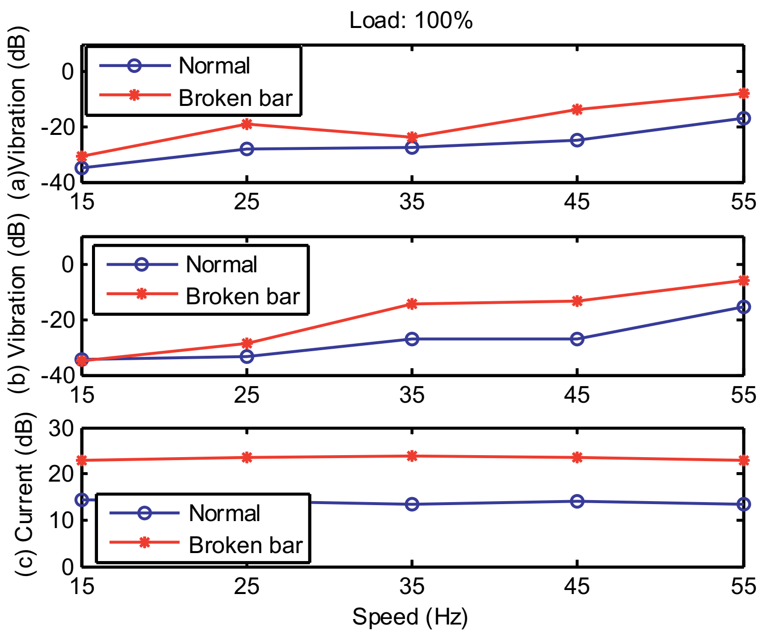


Figure 7. Influence of speed on the left sideband amplitudes at 100% load; (a) radial vibration; (b) axial vibration; (c) current.

Figure 7 compares the influence of speed on sideband amplitudes at full load conditions. The vibration sidebands in Figures 7(a) and (b) show that both the normal and faulty sideband amplitudes increase with speed. The comparison of current sideband amplitudes in Figure 7(c), on the other hand, shows that the sideband amplitudes and their differences remain almost constant of all speeds stated. The comparisons of sideband amplitudes at other load conditions show similar trends against the speed.

In general, for the detection of broken rotor bar faults, current analysis is more sensitive than vibration and acoustic methods. By analyzing the current signatures in the frequency domain, significant sideband differences could be observed. Vibration analysis is also able to show certain differences in sideband amplitudes. However, the differences are more likely to be influenced by shaft driving frequency than load. The acoustic analysis is barely able to detect broken rotor bar induced sideband differences due to high noise levels.

4.2 Bearing faults

The spectral comparison between normal and faulty bearings' radial vibration signatures at 25 Hz and 0% load condition is illustrated in Figure 8. The amplitudes of BPFO are 8.1 dB for the normal bearing and 33.1 dB for faulty bearing respectively, a difference of 25 dB. The amplitude difference of BPFI, about 23.8 dB, can also be observed around 123 Hz. On the other hand, the difference of BSF amplitudes (about 4 dB) is barely distinguishable. It is noted from Figure 8 that the radial vibration spectra contain a variety of spectral components and harmonics from shaft rotation and gearbox, etc. This may make the comparison of bearing conditions more difficult using vibration signatures. The spectral comparison of axial vibration signatures is similar to that of Figure 8.

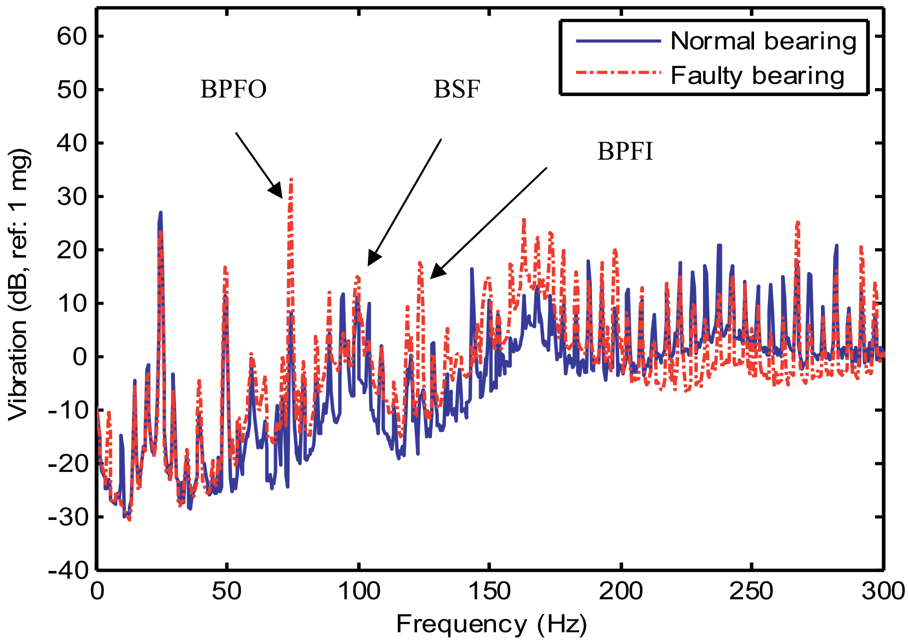


Figure 8. Spectral comparison of radial vibration signatures at 25 Hz and 0% load

Figure 9 shows the comparison of current spectra between normal and faulty bearings. The amplitude differences of bearing fault components are 3.1 dB (BPFO), -1.35 dB (BSF) and -2.7 dB (BPFI) respectively. Still, similar to the vibration spectra in Figure 8, the current spectra in Figure 9 are inundated with strong interferences and background noise.

The spectral comparison of acoustic signatures is shown in Figure 10. The comparison is made in two subplots because the sound pressure level distribution of the normal bearing is much higher than that of the faulty bearing. Such a difference in baseline spectra might be due to the acoustic contributions from background noise and other interferences. In Figure 10(b), the BPFO component at 76 Hz is clearly visible with an amplitude of 70.7 dB. Its second-order harmonic at 152 Hz has an amplitude of 57 dB. On the other hand, although the overall normal spectrum in Figure 10(a) is higher, it shows no sign of the BPFO component. The BPFI component at 123.2 Hz is barely observable with an amplitude of about 45 dB. The peak corresponding to the BSF could not be identified from the faulty spectrum in Figure 10(b).

From the above comparisons it can be seen that the BPFO fault is the most consistent one among the other faults. This is because the outer race fault has shorter distances to motor outer surfaces where the accelerometers are mounted. The rolling element fault, on the other hand, is not present at all test conditions as explained earlier (the faulty ball may change its rotational axis and thus may not always be over rolled at the same point). The BPFI fault is less strong in appearance than the BPFO fault as the BPFI related signal may undergo more attenuation through the bearing. Hence the following comparisons of the influences of speed and load on bearing faults are limited to outer race fault only.

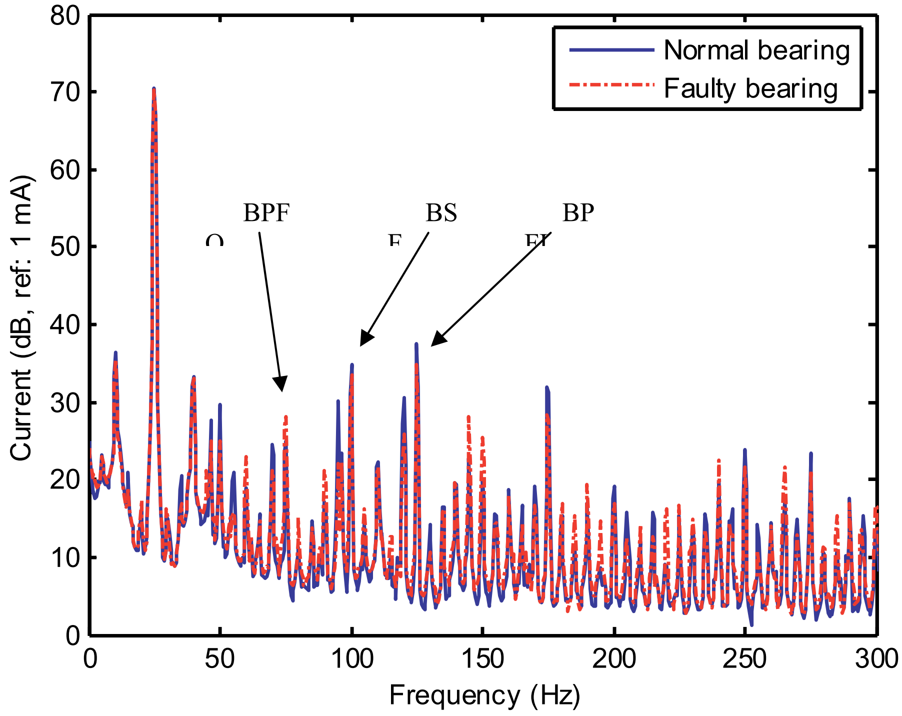


Figure 9. Spectral comparison of current signatures at 25 Hz and 0% load

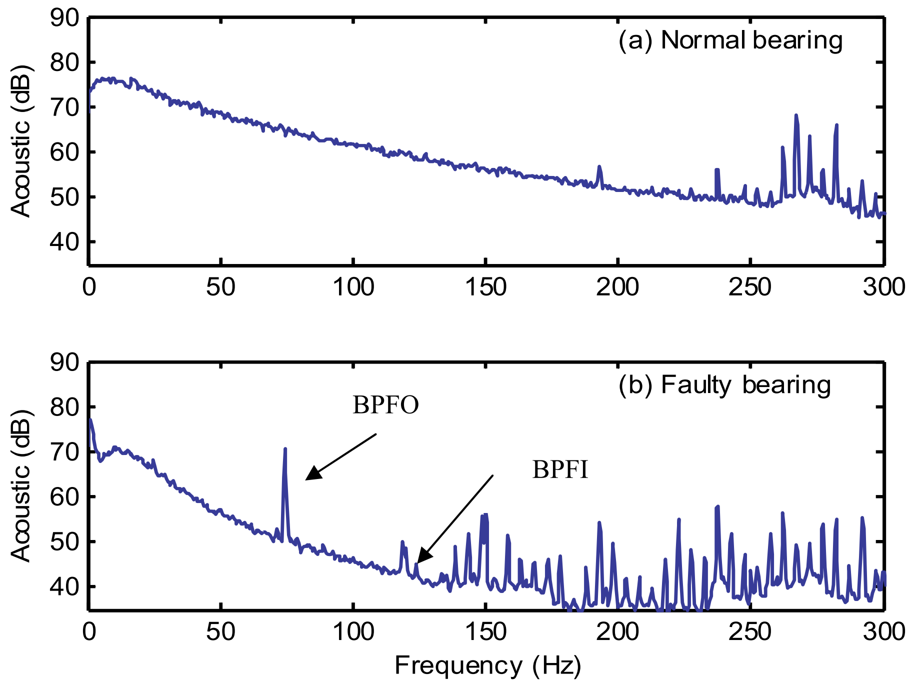


Figure 10. Spectral comparison of acoustic signatures at 25 Hz and 0% load

Figure 11 shows the influence of load on BPFO amplitudes at 15 Hz from the radial vibration signatures. First, both the normal and faulty BPFO spectral amplitudes decrease with the increase of load. Hence, the influence of load on bearing fault spectral amplitudes is limited. This is confirmed by the fact that the increase of load resulted in the decrease of spectral amplitudes at other driving frequencies and other faults (e.g. BSF and BPFI). Second, the amplitude differences between the normal bearing (circled line) and faulty bearing (stared line) in Figure 11 are the result of BPFO fault.

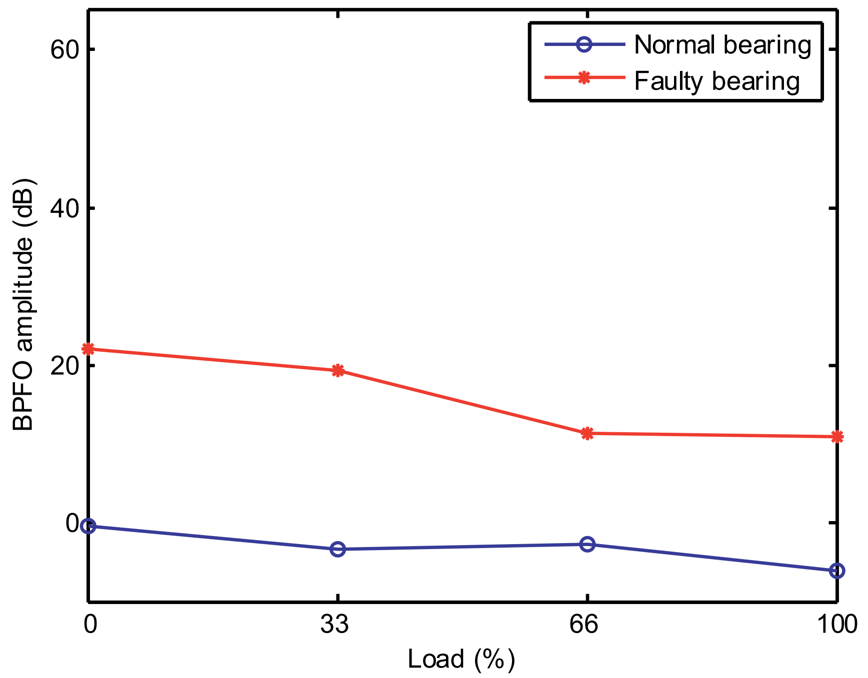


Figure 11. Influence of load on radial vibration BPFO amplitude at 15 Hz

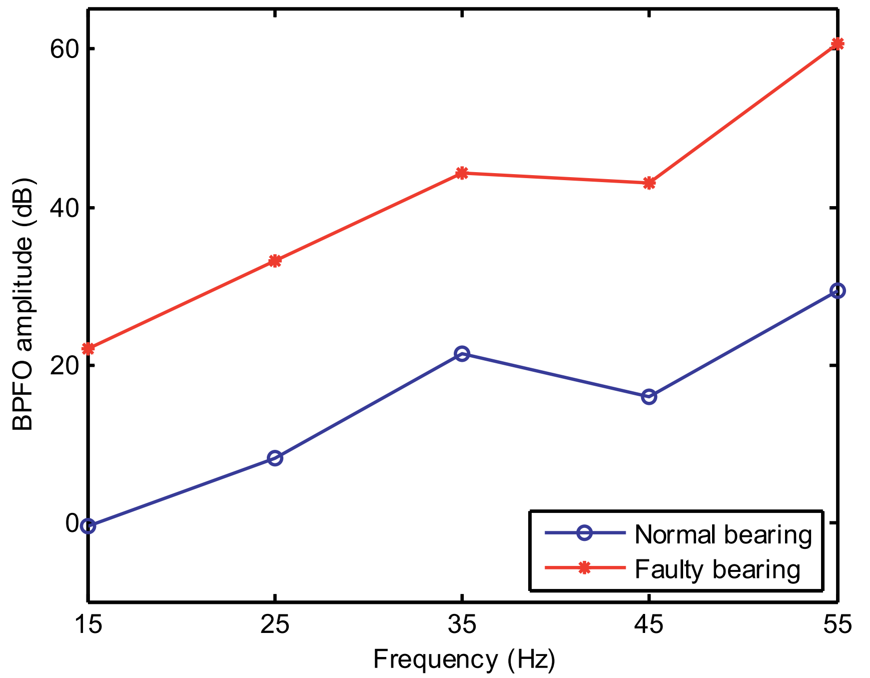


Figure 12. Influence of speed on radial vibration BPFO amplitude at 0% load

The influence of driving frequency on the radial vibration's BPFO spectral amplitudes is illustrated in Figure 12. In Figure 12, both the normal and faulty bearing spectral amplitudes increase with motor speed. The comparison clearly indicates that the spectral amplitudes are more prone to be influenced by motor speed. Figure 13 shows the acoustic spectral comparison of BPFO amplitudes between normal and faulty bearings at no load (0% load). Similar to vibration spectra in Figure 12, both amplitudes increase with motor speed.

In general, the analysis of vibration signatures is able to show significant differences in rolling element bearing faults over all load and speed conditions. However, vibration signatures contain a variety of interferences and harmonics from other vibration sources such as the gearbox. This, to some extent, makes the detection of bearing faults difficult in the presence of interferences. The analysis of acoustic signatures shows that the bearing fault frequencies are clearly identifiable in acoustic spectra. Furthermore, acoustic signatures contain fewer interferences and harmonics than vibration and current signatures in lower frequency bands. This makes acoustic analysis very attractive in the detection of bearing faults. The baseline spectral difference is clearly due to the change of motors. The current signature is least sensitive to bearing faults. Hence bearing faults would more likely induce significant changes in vibration and acoustic signatures rather than in current signatures.

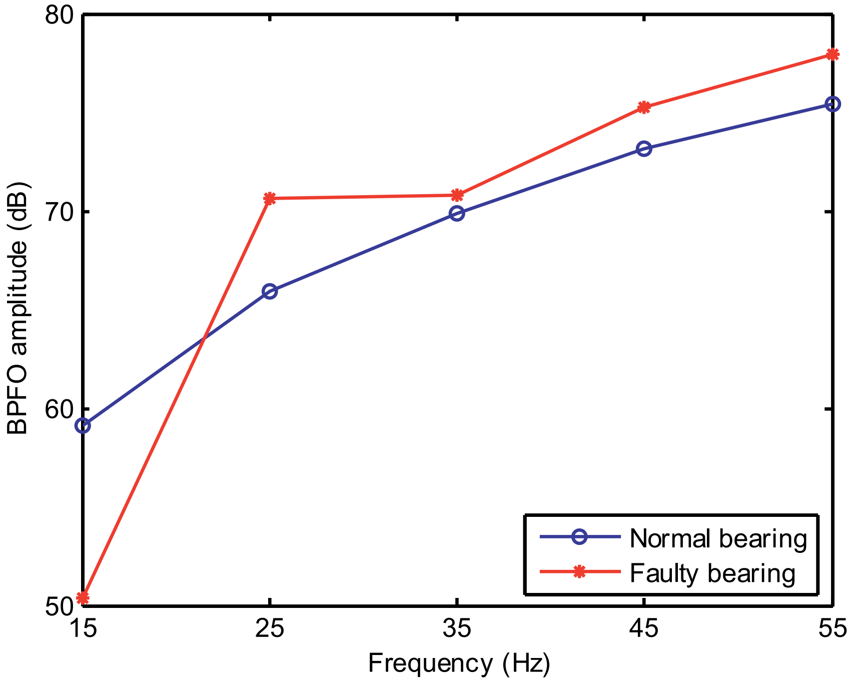


Figure 13. Influence of speed on acoustic BPFO amplitude at 0% load.

5 CONCLUSIONS

For the detection of the motor broken rotor bar fault, the stator current is the most sensitive method. The spectral analysis of stator current signatures indicated that the first few order sideband harmonics could be identified from the Fourier series based spectra. The amplitudes of these sidebands increased when the load was increased. The analysis of both the radial and axial vibration signatures was also able to show the spectral differences between motors. Contrary to the stator current method, the sideband amplitudes from the vibration signatures depended more on motor speed than on external load. The acoustic signatures also contain motor broken rotor bar information. The spectral analysis of the acoustic signatures was able to reveal the sidebands around the driving frequency.

Comparatively, the vibration method is most sensitive to the detection of bearing faults. The spectral differences between normal and faulty bearings at bearing fault frequencies are higher than those from current and acoustic spectra. However, vibration signatures contained a variety of harmonics and interferences making it difficult to identify bearing fault frequencies. Similar to vibration analysis, the current signatures were also inundated with harmonics and interferences make the method less sensitive to bearing faults. The analysis of bearing acoustic signatures showed that most bearing fault-related components could be easily identified in acoustic spectra. In the lower frequency band, the acoustic signatures contained fewer interferences and harmonics. This makes the acoustic analysis more attractive in bearing fault detection.

In general, the successful detection of induction motor faults depends on the selection of appropriate monitoring methods. From the results presented in this paper, it can be stated that the stator current method is sensitive to motor electrical faults while the vibration method is sensitive to motor mechanical faults. The acoustic analysis can be used as a supplementary method in the presence of noise and interferences.

6 ACKNOWLEDGEMENTS

The authors would like to thank the Natural Sciences and Engineering Research Council of Canada and Syncrude Canada Ltd for their financial support. The authors also wish to thank Lingxin Li for her assistance in collecting motor data.

7 REFERENCES

[1] Kliman GB, Koegl RA, Stein J, Endicott RD and Madden MW. (1988) Noninvasive detection of broken bars in operating induction motors. *IEEE Trans. Energy Conversion*, Vol. 3, No. 4, 873–879.

[2] Kryter RC and Haynes HD. (1989) Condition monitoring of machinery using motor current signature analysis. *Sound & Vibration Magazine*, 14-21.

[3] Benbouzid MEH. (2000) A review of induction motors signature analysis as a medium for faults detection. *IEEE Trans. Power Electronics*, Vol. 14, No. 5, 984-993.

[4] Betta G, Paolillo A and Pietrosanto A. (2002) A DSP-based FFT-analyzer for the fault diagnosis of rotating machine based on vibration analysis. *IEEE Trans. Instrumentation and Measurement*, Vol. 51, No. 6, 1316-1322.

[5] Liang B, Iwnicki SD and Ball AD. (2003) Asymmetrical stator and rotor faulty detection using vibration, phase current and transient speed analysis. *Mechanical Systems and Signal Processing*, Vol. 17, No. 4, 857-869.

[6] Benbouzid MEH and Kliman GB. (2003) What stator current processing-based technique to use for induction motor rotor faults diagnosis? *IEEE Trans. Energy Conversion*, Vol. 18, No. 2, 238-244.

[7] Benbouzid MEH, Nejjari H, Beguenane R and Vieira M. (1999b) Induction motor asymmetrical faults detection using advanced signal processing techniques. *IEEE Trans. Energy Conversion*, Vol. 14, No. 2, 147-152.

[8] Arthur N and Penman J. (2000) Induction machine condition monitoring with higher order spectra. *IEEE Trans. Industrial Electronics*, Vol. 47, No. 5, 1031-1041.

[9] [9] Yaz c B and Kliman GB. (1999) An adaptive statistical time–frequency method for detection of broken bars and bearing faults in motors using stator current. *IEEE Trans. Industry Applications*, Vol. 35, No. 2, 442-452.

[10] Filippetti F, Franceschini G, Tassoni C and Vas P. (1998) AI techniques in induction machines diagnosis including the speed ripple effect. *IEEE Trans. Industry Applications*, Vol. 34, No. 1, 98-108.

[11] Ye Z, Wu B and Sadeghian A. (2003) Current Signature Analysis of Induction Motor Mechanical Faults by Wavelet Packet Decomposition. *IEEE Trans. Industrial Electronics*, Vol. 50, No. 6, 1217-1228.

YOU CANNOT MANAGE WHAT YOU CANNOT MEASURE: AN INFORMATION SYSTEMS BASED ASSET MANAGEMENT PERSPECTIVE

Abrar Haider, Andy Koronios, Gerald Quirchmayr

School of Computer and Information Science, University of South Australia, Mawson Lakes, Adelaide 5095, Australia.

Abstract: Measuring the impact of implementation of information systems for asset management is a complex issue; due to the stochastic nature of process variables, substantial effects of information systems on the way users embrace these systems and consequently execute the business processes, and the high expectations that asset managing businesses associate with the use of information systems. This complexity can partly be attributed to the technology push strategy, rather than technology pull strategy, that asset managing businesses adopt to introduce information systems in to the business. Therefore, in order to take optimum advantage of information systems with regards to process efficiency, effective control, and management, it is important to have effective measurement mechanisms that help managers to measure Information systems utilisation for asset management process. This paper presents a measurement framework that assesses the impact of information systems at each stage of asset lifecycle management. The proposed framework based on generative learning, such that it examines the interpretation of asset management strategy through the use of information systems within the business and its assessment provides for strategic indicators for recalibration of asset management strategy as well as highlights the roadmap for future technology investments. This assessment allows for asset management processes and their stakeholders to adopt a technology pull strategy, which provides the strategic fit between processes and technology; thereby allowing the business to leverage optimum advantage of technology through rationalized investments.

Key Words: Asset Management, Asset Maintenance, Performance Evaluation, Information Systems.

1 INTRODUCTION

Rationalising size and quality of investments in information technologies along with improvements in organisational performance is on the strategic agenda of businesses around the globe. Nevertheless, businesses also take a few aspects for granted when it comes to adoption and implantation of information technologies. A common example is the massive investment in enterprise resource planning systems that carries the belief that its adoption would become a driver of operational excellence and an important source for business value. However, adoption of information technologies, particularly that of information systems (IS), entails taking stock of existing competencies, business architectures and processes that may be impacted by technology adoption [1]. This has particular relevance for the high risk and capital intensive businesses, such as engineering enterprises (see for example [2]). Generally, a fundamental issue with engineering enterprises is that they do not match their competencies, critical resources, and systems, with investments in technology. Consequently, there is little fit between the adopted technology and the way business is executed. This technology push strategy hampers the organisational adaptation to technology and the same time does not allow the business to take optimum advantage of technology. Nevertheless, in order to achieve this the business requires appropriate evaluation mechanisms that assess process performance, such that the assessment provides an understanding of the underperforming areas along with the reasons that contribute to the below par performance and the impact this underperformance man have on other areas of the business. This not only provides for the review of the performance of existing technology, but also provides for the useful indicators for future investments in technology. There are various advantages of having such a generative learning based performance measurement mechanism. Most significantly, it provides a holistic view of the whole business area and therefore highlights a road map for continuous improvement, whereby it facilitates the right choice of technology that best suits the business environment and ensures that the investments in technology are not just aimed at facilitating isolated business processes, rather each process contributes towards the overall strategic aims and objectives of the business.

This paper focuses on the IS utilised for asset life cycle management processes. It takes an asset lifecycle perspective and proposes a performance measurement framework for asset management, which assists engineering enterprises to evaluate the effectiveness of the IS utilised in the asset management lifecycle. The framework provides a cyclical approach to performance measurement such that it assesses and informs the role of IS in translating and informing the asset management strategy in a single cycle, thereby allowing for generative double loop strategic learning to enhance competitiveness of asset managing engineering enterprises.

2 PERFORMANCE ISSUES IN ASSET MANAGEMENT

An asset is the physical component of a manufacturing, production or service facility, which has value, enables services to be provided, and has an economic life of greater than twelve months [3], such as manufacturing plants, railway engines and carriages, aircrafts, water pumps, and oil and gas rigs. Accordingly, management of these assets represents a set of disciplines, methods, procedures and tools to optimise the whole life business impact of costs, performance and risk exposures associated with the availability, efficiency, quality, longevity and regulatory/safety/environmental compliance of a company's assets [4]. British Standards Institute [5] terms asset management as "combination of management, financial, engineering, and other practices applied to physical assets in pursuit of economic life-cycle costs. Its practice is concerned with specification and design for reliability and maintainability of plant machinery, equipment, buildings, and structures with their installation, commissioning, maintenance, modification, and replacement, and with feedback of information on design, performance, and costs". Nevertheless, the fundamental aim of asset management processes is the continuous availability of service, production and manufacturing provisions of assets. Consequently, asset management processes interact with a variety of other business processes, such as enterprise planning and product development, which necessitates timely and quality information interchange and coordination of activities with other functions within the business as well as with business partners, in order to allow for activities such as materials procurement, logistics, maintenance and repairs of assets, and customer relationship management. In crux, asset management is policy driven, information intensive, and is aimed at achieving cost effective peak asset performance. Asset management is not a set of isolated processes that support the life cycle of assets utilised by engineering enterprises, but these processes are actually embedded with the strategic processes of the business and have a direct impact on the overall profitability and efficiency of the businesses, as shown in Figure 1.

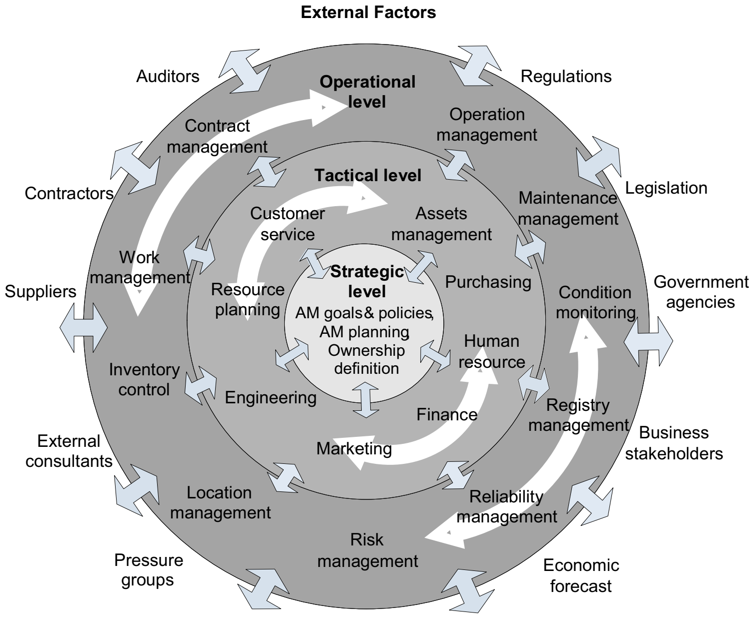


Figure 1: Asset Management Overview (source [3])

An asset lifecycle starts at the time of designing the manufacturing or production system, and typically illustrates, stages such as, asset commissioning, operation, maintenance, decommissioning and replacement. Market demand and supply dynamics derive product and services design, and this product and services design derives production. Production specifies the operational workload of an asset. Operational workload and asset design generate maintenance demands to keep the assets in running condition, whereas maintenance determines the future production capacity of the assets as well as their remnant life span. Table 1 summarises the processes and the activities involved in asset lifecycle management.

Table 1

Asset Management Lifecycle Management

Actions	Description
Strategic Planning	Planning of asset management objectives, functions, processes, technology, lifecycle costing, level of service, support infrastructure, and activities to meet business goals, stakeholder requirements, and demands management.
Focus	Strategic planning aims to provide the strategic fit between the overall business objectives and each activity performed in asset lifecycle management. Focus of strategic planning is to enhance competitiveness of the businesses through effective management of asset lifecycle.
Core Asset Management Actions	Creation or procurement, operation, condition assessment, maintenance, refurbishment, replacement, and disposal assets.
Focus	Ensuring availability, efficiency, quality, longevity and regulatory/safety/environmental compliance an asset at minimal costs
a. Asset creation/acquisition	Capital investment to acquire, construct, or improve an asset to satisfy or improve level of service, stakeholders' demand management, and production or manufacturing efficiency.
Focus	To ensure asset reliability, compliance and increased availability through effective design and best possible economic tradeoffs; for example high development costs might be traded off against lower maintenance costs through improved reliability. Development costs themselves may be offset through increased asset availability leading to increased production output.
b. Asset Operation	Smooth asset operation through close collaboration with maintenance function, and providing feedback on asset operational behavior to design function.
Focus	Efficiency and quality, regulatory/safety/environmental compliance of asset operation through conformance with planned operating conditions, as designed instructions, and environmental regulations.
c. Asset Operation Assessment	Asset operation condition monitoring to assess the ability of asset to meet required service levels and mitigate operational risks, including continuous or periodic inspection, assessment, measurement, reporting and interpretation of resulting information to indicate the condition of the asset in order to specify the nature and timing of maintenance.
Focus	To ensure asset longevity by proactively assessing asset condition in order to predict developing failure conditions, such that the subsequent maintenance execution and resources could be planned.
d. Asset Maintenance	Sustain and where necessary performance necessary repair and maintenance on assets in such a way that they continue to fulfil their functions and make the required value adding contribution to the manufacturing or production process, including actions such as spare supply chain management, maintenance planning and maintenance workflow execution, routine repairs, and testing.
Focus	Ensuring efficiency and longevity of asset by taking necessary actions to retain an asset to near original asset configuration and condition by decreasing its rate of deterioration, thereby minimising asset downtime.
e. Asset Rehabilitation	Rehabilitation of an asset through technology refresh, end of need, or maintenance; including actions such as treatments for improving condition either to attain as designed/as constructed condition or to exceed the as designed/as constructed condition in order for the asset to function at a satisfactory level of service for a prolonged time frame.
Focus	Review of asset configuration, type, location, and service delivery aspects, and their fit with the strategic business objectives.
f. Asset Replacement/Disposal	Investment aimed at replacing or reconstructing an existing asset in order to enhance level of service, and/or improved configuration, and/or change in physical location. Asset disposal or retirement means rationalisation of an asset according to safety/regulator requirements when it is no longer in service.
Focus	Continuous improvement opportunities in asset design and configuration, in order to provide elevated level of service.

Continued

Actions	Description
Asset Lifecycle Support	Asset lifecycle support is aimed at having requisite infrastructure supporting life cycle decisions of an asset, whereby the essential gaol is to identify all available options and select the most efficient and cost effective option.
Focus	Financial and non financial infrastructure aimed at enhancing the quality of asset lifecycle management.
Asset Lifecycle Costing	Calculation of asset lifecycle cost benefit analysis including asset design, asset procurement and/or construction, asset operation, asset maintenance and support, and asset retirement and disposal costs. It also contributes to assessment of high-cost contributors; cause and effect relationships; potential areas of risk; and identification of areas for cost reduction.
Focus	A systematic approach to recording and analysing costs associated with each stage of asset life cycle, so as to provide asset lifecycle management decision support in terms of economic tradeoffs and cost benefit analyses.
Asset Lifecycle Decision Support	Availability of complete, current, and accurate information on each asset, including its design, construction, configuration, location, workload, health status and history, cost benefit analysis, and the physical environment that it operates in.
Focus	An integrated approach to economic and performance tradeoffs and lifecycle decisions, through organization of information relating to asset design, operation, condition, maintenance, refurbishment; skills and support infrastructure to ensure asset reliability, availability, and quality; analytic models that predict future changes in asset design, operation, condition, as well as the variations in support mechanisms to forecast and plan for resources; and cost benefit analysis for tradeoffs regarding asset design, operation, maintenance, renewal, and decommissioning etc.
Asset Lifecycle Technology Support	Combination of hardware; software; information acquisition, processing, communication and storage infrastructure; and skills of employees to process information with the hardware and software; to accomplish various asset management processes and providing required outputs for effective asset management.
Focus	Strategic fit between technology and asset management processes, aimed at availability of timely, consistent, complete, accurate, and reliable information.
Asset Management Performance Assessment	Review and assessment of asset lifecycle management actions to measure their effectiveness in satisfying business needs.
Focus	Audit and assessment of implementation and execution of planned activities against documented gaols, objectives, strategies, and stakeholder requirements aimed at continuous improvement of the asset management process.
Asset Lifecycle Learning	Profiling asset operation, managing lifecycle knowledge, and feedback for cost effective, better understanding, and continuous improvements in asset design, operation, maintenance, reinvestment, compliance, and asset lifecycle support infrastructure and resources.
Focus	Preservation and use/reuse of asset management knowledge for subsequent strategic planning cycles.

(Adopted from [3], [4], [6-10])

It is clear from the above that IS utilised in asset management not only have to provide for the decentralized control of maintenance tasks but also have to act as instruments for decision support. An important aspect of asset lifecycle management is the learning or knowledge gained at each stage, which provides for the feedback to other processes. For example, asset operation profiling has significance for asset redesign as well as asset maintenance and asset operation cost benefit analyses and lifecycle decision support. Therefore, the most crucial feature of IS for asset management is their ability to provide an integrated view of asset specific information, such that engineering enterprises are able to leverage the full potential of the IS utilised for asset management. This requires appropriate hardware and software applications; quality, standardised, and interoperable information; appropriate skill set of employees to process information; and the strategic fit between the asset management processes and the IS.

3 INFORMATION SYSTEMS FOR ASSET MANAGEMENT

Information technologies are utilised at three levels for asset lifecycle management; firstly, they are used in assets, such as microcontroller operated motors. Secondly, information technologies that are used in facilitating asset lifecycle management support processes, such as digital sensors and other condition monitoring equipment, and process control infrastructure, such as Supervisory Control And Data Acquisition system (SCADA). Thirdly, the IS that not only facilitate execution of each stage of asset lifecycle but also provide for decision support for asset life cycle management, such as asset design, maintenance planning and execution, resource allocation, and maintenance workflow execution. Standardised and quality information, therefore, is the most important ingredient of effective asset lifecycle management.

Having its origin in mass production aimed at capturing market share, quality management calls for standardization of processes managed by standardised data and facts that focus on certain targets set by informed choices. However, value of these informed choices cannot be guaranteed in business areas where such positivist assumptions are invalid (see [11], [12]). Asset maintenance is one such area within asset management, which has by and large routine processes, yet these processes operate in unpredictable environments and function on basically non-market transactions. Hence, the basic issue in asset management IS is not just the quality of converting input to output, but also the control of information guiding it and the eventual use and reuse of the resulting knowledge for decision support and enterprise wide planning. In simple words, the issue is not only doing things right, but also to have knowledge of what are the right things to do. On the contrary, while maintenance activities have been carried out ever since advent of manufacturing; modelling of an all inclusive and efficient maintenance system has yet to come to fruition ([13], [14]). This is mainly due to continuously changing and increasing complexity of asset equipment, and the stochastic nature or the unpredictability of the environment in which assets operate, along with the difficulty to quantify the output of the maintenance process itself [15]. At the same time, in response to the competitive pressures, maintenance strategies that once were run-to-failure are now fast changing to being condition based, thereby necessitating integration of asset management decision systems and computerized maintenance management systems in order to provide support for maintenance scheduling, maintenance workflow management, inventory management, and purchasing [16]. Businesses are now aiming for ways to providing direct connections from their asset management systems to Maintenance, Repair, and Overhaul/Operations procurement systems, which may allow for paperless purchasing of parts and offer considerable time and cost savings compared with traditional purchasing methods.

Nevertheless, in order to leverage the advantages of technology asset managing organisations need to find the fit between technology and the processes. For example, a typical water pump station in Australia is located away from major infrastructure and has considerable length of pipe line that brings water from the source to the destination. In this situation, assets are deployed over an area of various kilometres; however, the demand for water supply is continuous for twenty four hours a day, seven days a week. Although, the station may have some kind of an early warning or process control and condition monitoring system installed, such as Supervisory Control and Data Acquisition (SCADA), maintenance labour at the water stations and along the pipeline is limited and spares inventory is generally not held at each station. Therefore, it is imperative to continuously monitor asset operation (which in this case constitutes equipment on the water station as well as the pipeline) in order to sense asset failures as soon as possible and preferably in their development stage. However, early fault detection is not of much use if it is not backed up with the ready availability of spares and maintenance expertise. The expectations placed on water station by its stakeholders are not just of continuous availability of operational assets, but also of the efficiency and reliability of support processes. Elimination and control of production irregularities and disturbances is, therefore, necessary for continuous production and service provision, and customer satisfaction. However, situation on ground is far from being perfect, for example, data is captured both electronically and manually, in a variety of formats, shared among an assortment of off the shelf and customized operational and administrative systems, communicated through a range of sources and to an array of business partners and sub contractors; and consequently any disparity in data at the acquisition level leads to the inability of quality decision support for asset lifecycle management [17]. In the prevailing circumstances, existing asset management IS could best be described as useful pools of isolated data that are not being put to effective use to create information and business intelligence. Most engineering enterprises mature technologically along the continuum of standalone technologies to integrated systems, and in so doing aim to achieve the maturity of processes enabled by these technologies and the skills associated with their operation [1]. Konradt et al [18] further assert that engineering enterprises adopt a traditional technology-centred approach to asset management, where technical aspects command most resources and are considered first in the planning and design stage. Processes, human and other organisational factors are only considered relatively late in the process, and sometimes only after the systems are operational. Consequently, engineering enterprises struggle to implement cost effective asset management strategies that best suit the business; develop lifecycle asset management competencies; plan an effective exit strategy for assets rendered obsolete through technology refresh or through end of need; and provide a credible charge-back system to allocate costs to the business lines and thus ensure that everyone is involved in avoiding redundancy and wastage of efforts [1]. In order to make an effective investment into technology, it is important for the business to take stock of the existing IS and the asset management processes that they enable.

This highlights the need for a comprehensive performance measurement system that not only provides insights into the effectiveness of the IS used for asset lifecycle management, but also provide feedback on their fit with other business information systems so as to provide a lifecycle perspective of asset management to asset managers. Such a performance measurement system needs to be three dimensional (as shown in Figure 2), with each asset lifecycle management process being measured against the four dimensions of an IS, i.e. information, staff skills, software, and hardware; as well as the impact of each asset lifecycle process in terms of asset efficiency, availability, longevity, quality and reliability, and compliance with the environmental and operating legislations.

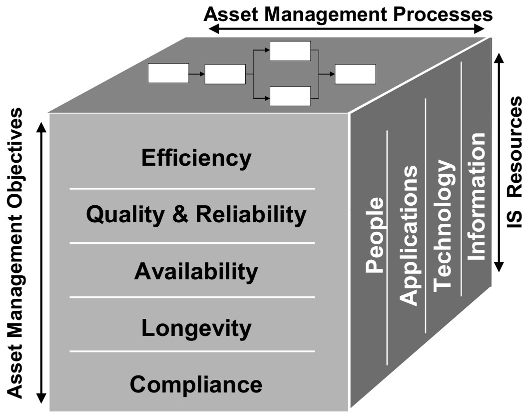


Figure 2: Measurement dimensions for IS based Asset Management

4 PERFORMANCE MEASUREMENT SYSTEMS

There has been numerous business improvement methodologies developed and implemented. These methodologies represent a blend of theory and practice, with each having its own way of performance measurement that is largely dependent upon the target business area. Some of the leading initiatives in this regard include Benchmarking, total quality management, Six Sigma, European Foundation for Quality Management Business Excellence Model (EFQM), business process re-engineering, and balanced scorecard. These methodologies constitute the basis of the most of performance measurement and management initiatives tailored by businesses to meet their needs. Remenyi et al [19] summarise the methodologies developed for performance assessment, and suggest that their focus has been on

a. Strategic match analysis and evaluation,

b. Value chain assessment (organisation and industry),

c. Relative competitive performance,

d. Proportion of management vision achieved,

e. Work-study assessment,

f. Economic assessment - I/O analysis,

g. Financial cost benefit analysis,

h. User attitudes,

j. User utility assessment,

k. Value added analysis,

l. Return on management, and

m. Multi-objective, multi-criteria methods.

Engineering enterprises have adopted these methodologies in a variety of ways aimed at different business areas and processes, such as for manufacturing planning and control [20], product development process [21], human resources development ([22], [23]) and service or facility management [24]. Although research has paved the way for major developments in the filed of business improvement, yet it is interesting to note that asset performance measurement has been largely limited to physical inspection of plant and equipment for its health assessment. Nevertheless, role of IS in asset management is not just of a process enabler, but also of translating strategic business goals and objectives into action. Therefore, it is important to have a performance measurement mechanism that provides for assessment of this strategic translation and also highlights the gaps such that corrective actions could be taken to enhance the competitive of the overall business. These characteristics entail that an appropriate performance measurement system for asset management should:

a. Focus on business processes as well as the structures that deliver value [25];

b. Integrate different aspects of asset management, such that they constitute a chain for business competitiveness ([26], [27]);

c. balance the needs of various stakeholders, such as business partners or third party service providers, customers, employees, regulatory agencies, and society at large ([20]; [25]);

d. be information driven such that it provides inputs to strategy re-calibration rather than being steered by the business strategy alone ([28], [29], [25]);

e. Conform to business objectives ([[28], [30]);

f. Aim at competency development and business intelligence infrastructure development in order to create and sustain value for asset management processes ([27], [25]; and

g. Provide financial [27] as well as non financial assessments [25].

In engineering enterprises strategy is often built around two principles competitive concerns and decision concerns. Competitive concerns set the gaols of manufacturing, whereas decision concerns deal with the way these goals are to be met [48]. It is therefore, important to assess the role of IS in providing strategic executive decision support for value added asset management, in terms of the choice of assets, their demand management, support infrastructure to ensure smooth manufacturing or service provision, and efficient ways of doing business. These efficient ways mean choices of the business such as business partners, outsourcing of asset management functions, and service provision to third parties [47]. Fundamental requirement of IS at this stage is to provide decision support in terms of indications on economic tradeoffs and/or alternatives that help senior management in making strategic decisions. These decisions are not made in isolation; in fact, they are made in conjunction with the asset demand and stakeholders requirements. Therefore at the strategic level it is important to assess the effectiveness and ability of the IS in terms of enabling strategic asset management planning, goals, objectives and executive decisions.

A usual manufacturing cycle starts with specification of the products and services that the business aspires to offer its customers in conformance with its business strategy. Any mismatch between what the market demands, manufacturing process design, and planning has a detrimental effect on the overall performance of the business [49]. Nevertheless, this specification illustrates the types of assets and processes that the business needs to put in place to produce services and products, also called asset demand specification. At this stage, it is important to assess if the designers have access to integrated information in order to process it to obtain knowledge of factors such as, the characteristics of the environment that the business operates in; design, configuration, and workload of each asset; asset and/or site layout design and schematic diagrams/drawings; asset bill of materials; analysis of maintainability and reliability design requirements (e.g. HAZOP); and failure modes, effects and criticality identification for each asset. Planning choices at this stage drives asset behaviour, therefore it is important to assess if the existing information systems are capable of providing the right information, such that choices could be made to ensure asset availability, reliability and quality of its operation, and efficiency.

Design of an asset has a direct impact on its productivity. Productivity, itself, is concerned with minimising the disturbances relating to production or service provision of an asset. A production or service provision disturbance is an unplanned or undesirable function or failure of an asset [55]. It can be classified as asset downtime, speed or operation, and quality losses. It is important to note that disturbances do not only occur due to a mechanical or electrical fault, they can also occur due to some process and procedural issues. Disturbances occur in one of the three ways, as suggested by El-Haram [56], i.e., when an asset become inoperable suddenly, and can no longer perform its required operations; when an asset cannot fulfil some or all of its operations at the same performance standard as originally specified; or when an asset gradually deteriorates to an unsatisfactory level of performance or condition, and its continued operation is unsafe, uneconomical or aesthetically unacceptable. These disturbances can further be attributed to many causes. According to one study, more than one third of the production disturbances were caused by design errors [57]. Therefore, at this stage it is important to assess if the IS are capable of providing integrated feedback to maintenance and design functions regarding factors such as asset performance profiling; detection of manufacturing or production process defects; asset condition monitoring and inspection; asset failure notifications; quality control and assurance at each stage of asset operation; and processes, tests, analysis (FMECA etc) and standards to be applied for the validation of asset performance.

Every asset generates some maintenance during its lifecycle. The aim of asset lifecycle support is to keep or restore and asset to its original or near original condition and service levels. This requires integration of technical, administrative, and operational information of asset lifecycle, such that informed and cost effective choices could be made about maintaining an asset and its remnant lifecycle. Therefore, the IS should be capable of handling the maintenance workflow execution as well as project management along with processing and providing information on factors such as, Asset failure and wear pattern; routine maintenance work plan generation; emergency maintenance scheduling and follow up actions (such as disaster containment, maintenance resources allocation etc); asset shutdown scheduling; maintenance simulation; testing after servicing/repair treatment; identification of asset design weaknesses; and asset operation cost benefit analysis. Maintenance, however, influences many areas of the business, such as asset availability in supporting just in time manufacturing principles [50], relationship between technology and operations [58], product quality [7], and achieving and sustaining a safe workplace and environment [38]. Therefore, it is important to assess the role of IS in providing for the up-to-date information on each asset's health to other functions such as asset design and asset operation. This information is extremely important in asset improvements/refurbishments as well as in finding optimal ways of operating an asset.

Nevertheless, an important aspect of effectiveness of asset management initiatives is the efficiency with which different processes are executed and the competencies of employees who carry out these processes. This efficiency comes from a variety of factors such as application of appropriate socio-technical systems principles and techniques make asset management creative and more productive; availability of asset lifecycle information and knowledge throughout the organisation; and development of employees skills to effectively use technologies associated with asset lifecycle management. IS provide for this value chain integration of asset management, therefore, it is important to assess the effectiveness of IS in terms of operational efficiency, and the ability of employees to operate IS and process information contained in them.

A common trend among engineering enterprises is the outsourcing of core activities such as maintenance. This trend is quite common among for complex assets, such as aircrafts, and oil and gas rigs. In these circumstances neither the asset owning business, and nor the asset maintaining business has a complete understanding of asset behaviour, which obviously impacts asset lifecycle decision support. It is important to assess the level of integration that the IS provide in bringing together business stakeholders, such as business partners, customers, and regulatory agencies like environmental and government organisations. The idea behind this is to assess the level of information and knowledge being shared among the stakeholders to enhance the efficiency and competencies of the asset management process, by examining the IS maturity for factors such as third party services management; contract management; integration of inter organisational applications and processes; customer feedback and satisfaction measurement; and sharing common coding and databases with business partners.

5 IS BASED ASSET MANAGEMENT EVALUATION

From the discussion above, two characteristics stand out. First, IS based asset management performance evaluation entails taking stock of effectiveness of IS in facilitating each asset management lifecycle process; and second, how the IS provide for the functional integration between various asset management processes. Table 2 illustrates an evaluation framework for IS utilised for asset management. In this framework asset management process has been decomposed into seven perspectives, which cover the whole of asset lifecycle as well as provide for the functional interrelations; thereby forming a value chain that provides value added asset management.

Table 2

IS based Asset Management Performance Measurement Perspectives

Perspective	Description	References
Design	Planning of design and/or improvements in asset design and manufacturing processes according to feedback from asset installation, commissioning, operation, maintenance and replacement; and aimed at demand management as per stakeholders' expectations.	[3], [14], [32], [33], [34]
Productivity	Asset availability and reliability by managing quality of operations, and mitigating risks posed to assets and to their operating environment; ensuing flow of information to maintenance and design regarding asset operation feedback.	[3], [4], [16], [20], [33], [34], [35], [37]
Support	Financial and non financial resources support for asset lifecycle aimed at enhancing the quality of asset operation; decision support for effective asset health management; integration of technical, administrative, and operational information of asset lifecycle aimed at keep it, or restore it to original or near original condition and service levels.	[3], [6], [7], [8], [14], [20], [32], [36], [38], [39], [44], [50]
Operational Efficiency	Measure of the factors that contribute to the overall effectiveness and integration of asset lifecycle processes and functions; application of appropriate socio-technical systems principles and techniques make asset management creative and more productive; business value chain integration; skills development of employees to effectively use technologies associated with asset lifecycle management; availability of asset lifecycle information and knowledge throughout the organisation.	[3], [4], [7], [13], [16], [34], [36], [40], [41], [42], [43], [44]
Stakeholders	Integration of business stakeholders to achieve higher levels of asset management through communication and collaboration, sharing of information on asset lifecycle with subcontractors and business partners, and compliance to regulatory and environmental regulations.	[2], [3], [20], [37], [38], [45], [46]
Competitiveness	Availability of executive decision support indicators such as, asset lifecycle cost benefit analysis, operation profile, asset register to better manage asset demand and needs of stakeholders ; ability to measure the actual performance of asset lifecycle management processes against planned gaols and objectives to enhance competitiveness of the business.	[3], [6], [7], [8], [9], [14], [20], [33], [38], [47], [48], [49]
Learning	Profiling asset lifecycle, and managing lifecycle knowledge for better understanding of/and improvements in asset design, operation, maintenance, reinvestment, and compliance.	[3], [6], [7], [8], [10], [43], [51], [52], [53], [54]

Figure 3 provides the graphical representation of the evaluation framework. Here IS act as the bonding glue that provides for the functional integration of asset management process. This framework provides double loop generative learning, as it allows for assessment of IS in terms of translation of strategy into action; and also highlights the gaps between the existing and desired levels of performance, thereby informing the business strategy in the same cycle. In so doing, the organisation learns from one cycle, applies the learnings to the same paradigm and assesses its results in the next cycle. At the same time highlighting of the underperforming areas creates the need for investment in technology, which not only allows for the right choice of technology but also helps in its acceptance in the organisation. This framework is learning centric that provides for preserving the asset lifecycle knowledge, such that it is accessible to every function within the organization. This learning perspective illustrates assessing the way IS are utilised in an engineering enterprise to preserve the knowledge that it creates through other perspectives regarding asset performance and operational knowledge; asset health history; asset operation cost benefit analysis; and historic asset lifecycle performance information. IS evaluation at this perspective means assessing the ability of the existing IS to provide multi dimensional analyses on asset lifecycle information, such that asset managers can take better informed decisions about the asset lifecycle and asset management processes. These analyses provide triggers for change regarding asset design, operation, maintenance, risk management, and other aspects of asset lifecycle management, thereby allowing for creativity in asset lifecycle management. In so doing, this framework provides for continuous improvement based on generative learning.

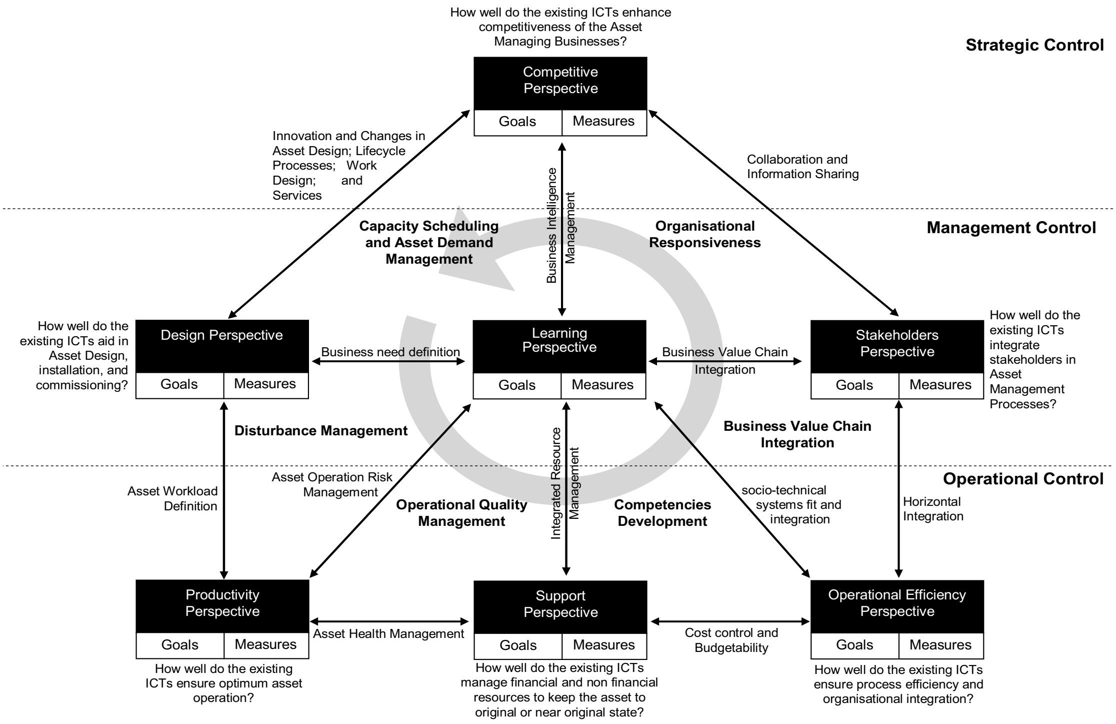


Figure 3: IS Based Asset Management Performance Evaluation Framework

6 DATA COLLECTION

In order to provide an integrated assessment of IS based asset management, as suggested in figure 2, an assessment approach as illustrated in Table 3 could be adopted. Processes under each perspective could be assessed for dimensions of IS, i.e. information, people, software, and technology on a scale of 1 to 5. In the same way, depending upon the characteristics of the process, its purpose, i.e. primary or secondary, could be assessed. This information could be collected though surveys and with the help of Analytic Hierarchy Process (AHP) (Saaty 1990) and Multi-Attribute Utility Theory (MAUT) (Goicoechea et al. 1982) it could be aggregated to provide performance measurement of IS for processes relating to each of the perspectives identified for asset management process.

Table 3

Data Collection Tool for the Proposed Framework

	Asset Performance Criteria						IS Dimensions			
	Efficiency	Longevity	Availability	Compliance	Reliability		People	Software	Technology	Information
Design Perspective										
Asset bill of materials										
Naming and numbering of assets for configuration management										
Use of test and historical maintenance data of similar assets in establishing and evaluating operational tasks										
Identification of risks posed to asset operation (environmental, OH&S)										
Analysis of maintainability and reliability design requirements (e.g. HAZOP)										
Failure modes, effects and criticality identification										
………………………………..										
Productivity Perspective										
Processes										
Support Perspective										
Processes										
Operational Efficiency Perspective										
Processes										
Stakeholders Perspective										
Processes										
Competitiveness Perspective										
Processes										
Learning Perspective										
Processes										

7 CONCLUSION

This research provides the basis for research into comparative analysis of IS based asset management for industrial and infrastructure assets to be conducted through the Cooperative Research Centre for Integrated Engineering Asset Management (CIEAM). This paper has proposed and theocratically demonstrated an approach for linking asset management to strategic competitiveness of an engineering enterprise through IS maturity assessment and control. It has particularly emphasised the use of IS for functional integration by providing a strategic fit between the structure and infrastructure of asset lifecycle management process. It has also shown how asset managing businesses could benefit by taking a lifecycle perspective of asset management, such that assets are treated as business enablers rather than just production or service provision enablers.

8 REFERENCES

[1] 1. Haider, A, and Koronios, A., (2005), "ICT Based Asset Management Framework", International Conference on Enterprise Information Systems (ICEIS), 3, pp. 312-322.

[2] 2. Liyanage, J.P., Kumar, U., (2000), Utility of maintenance performance indicators in consolidating technical and operational health beyond the regulatory compliance, Doerr, W.W., Safety Engineering and Risk Analysis: The International Mechanical Engineering Congress and Exposition-2000, pp.153-160.

[3] 3. IIM, (2002), International Infrastructure Manual, National Asset Management Steering Group, Australia New Zealand Edition, Thames, ISBN 0-473-09137-2

[4] Woodhouse, J., (2001), “Asset Management”, The Woodhouse Partnership Ltd, online accessed on September 10, 2005 at http://www.plant-maintenance.com/articles/AMbasicintro.pdf

[5] British Standard BS 3811:1993 (1993) Glossary of Terms Used in Terotechnology, British Standards Institution Publishing Limited. ISBN: 0 580 22484 8 .

[6] Moubray, J. M., (2003), 21st Century Maintenance Organization Part I: The Asset Management Model, Maintenance Technology, accessed online at http://www.mt-online.com/articles/0203_asset_mgmtt.cfm, on July 12, 2005.

[7] Moubray, J.M., (2000), Reliability-centred Maintenance, Butterworth-Heinemann, Oxford.

[8] Blanchard, B. S., (1997) Logistics Engineering and Management, 6th Edition, Prentice-Hall, Upper Saddle River, NJ.

[9] Husband, T.M (1976) Maintenance Management and Terotechnology, Farnborough : Saxon House, England.

[10] Argyris, C., (1977), Double Loop Learning in Organizations, Harvard Business Review. 55, September-October, pp.115-125.

[11] Kirkpatrick, I., and Lucio, M. M., (1995), “The Politics of Quality in the Public Sector”, Routledge, London.

[12] Adams, R., (1998), “Quality Social Work”, Macmillan, London.

[13] Duffuaa, S., Ben-Daya, M., Al-Sultan, K., and Andijani, A., (2001), A Generic Conceptual Simulation Model For Maintenance Systems, Journal of Quality in Maintenance Engineering, 7(3), pp. 207-219.

[14] Yamashina, H., (2000), Challenge to world class manufacturing, International Journal of Quality & Reliability Management, 17(2), pp. 132-143.

[15] Duffuaa, S.O., Cambel, J.D., and Raouf, A., (1999), “Planning and Control of Maintenance Systems: Modelling and Analysis”, John Wiley & Sons, New York.

[16] Bever, K., (2000), Understanding Plant Asset Management Systems, Maintenance Technology, July/August, pp. 20-25

[17] Haider, A., Koronios, A., (2003), “Managing Engineering Assets: A Knowledge Based Approach through Information Quality, Proceedings of 2003 International Business Information Management Conference, pp.443-452.

[18] Konradt, U., Zimolong, B., and Majonica, B. (1998), User-Centred Software Development: Methodology and Usability Issues, The Occupational Ergonomics Handbook, Karwowski, W., and Marras, W.S., (eds). USA, CRC Press.

[19] Remenyi, D., Money, A., Sherwood-Smith, M., Irani, Z., (2001) Effective Measurement and Management of IT Costs and Benefits, Second Edition, Butterworth-Heinemann, NY.

[20] Tangen, S., (2004), Performance measurement: from philosophy to practice, International Journal of Productivity and Performance Management, 53(8), pp. 726-737

[21] O'Donnel, F., Duffy, A.H.B., (2002), Modelling design development performance, International Journal of Operations & Production Management, 22(11), pp.1198-1221.

[22] Kelly, J., Gennard, J., (2001), Power and Influence in the Boardroom, Routledge, London.

[23] Gibb, S., (2002), Learning and Development: Process, Practices and Perspectives at Work, Palgrave, London.

[24] Wilson, R. C., (2000), ISO 14000 Insight: Benefits of Implementing an EMS, Pollution Engineering, 32(2), pp-33.

[25] Neely, A., Mills, J., Gregory, M., Richards, H., Platts, K., Bourne, M. (1996), Getting the Measure of Your Business. Findlay, London.

[26] Suwignjo, P., Bititci, U.S., Carrie, A.S., (2000), Quantitative models for performance measurement systems, International Journal of Production Economics, 64, pp.231-241.

[27] Neely, A., Adams, C., (2001), The performance prism perspective, Journal of Cost Management, 15(1), pp.7-15.

[28] Kaplan, R.S., Norton, D.P., (1996), The Balanced Scorecard: Translating Strategy into Action, Harvard Business School Press, Boston, MA.

[29] Bititci, U.S., (2000), Dynamics of performance measurement systems, International Journal of Operations & Production Management, 20(6), 692-704.

[30] Bititci, U.S., Mendibil, K., Martinez, V., and Albores, P., (2005), Measuring and managing performance in extended enterprises, International Journal of Operations & Production Management, 25(4), pp. 333-353.

[31] Kaplan, R.S., Norton, D.P., (2000), The Strategy Focused Organisation: How Balanced Scorecard Companies thrive in the New business Environment - Measures that Drive Performance, Harvard Business School Press, Boston, MA.

[32] Jonsson, P., and Mattsson, S., (2003), The implications of fit between planning environments and manufacturing planning and control methods, International Journal of Operations & Production Management, 23(8), pp. 872-900.

[33] Raouf, A. S. I., (2004), Productivity enhancement using safety and maintenance integration: An overview, The International Journal of Systems & Cybernetics, 33(7), pp. 1116-1126.

[34] Fernando, J. B. L., and de Carvalho, M. M., (2005), Changing product development process through information technology: a Brazilian case, Journal of Manufacturing Technology Management, 16(3), pp. 312-327.

[35] Suzuki, T., (1994), TPM in Process Industries, Productivity Press, Cambridge, MA

[36] Mathew, S., (2004), Optimal inspection frequency: A tool for maintenance planning/forecasting, International Journal of Quality & Reliability Management, 21(7), pp. 763-771.

[37] Seth, D., and Tripathi, D., (2005), Relationship between TQM and TPM implementation factors and business performance of manufacturing industry in Indian context, International Journal of Quality & Reliability Management, 22(3), pp. 256-277.

[38] Zutshi, A., and Sohal, A. S., (2005), Integrated management system: The experiences of three Australian organisations, Journal of Manufacturing Technology Management, 16(2), pp. 211-232.

[39] Oke, S. A., (2005), An analytical model for the optimisation of maintenance profitability, International Journal of Productivity and Performance Management, 54(2), pp. 113-136.

[40] Al-Najjar, B., (1996), Total quality maintenance An approach for continuous reduction in cost of quality products, Journal of Quality in Maintenance Engineering, 2(3), pp. 4-23.

[41] Blanchard, B. S., (1997), An enhanced approach for implementing total productive maintenance in the manufacturing environment, Journal of Quality in Maintenance Engineering, 3(2), pp. 69-80.

[42] Cassady, C., Pohl, E., and Murdock, W., (2001), Selective Maintenance Modelling For Industrial Systems, Journal of Quality in Maintenance Engineering, 7(2), pp. 104-117.

[43] Argyris, C. (1993) Knowledge for Action. A guide to overcoming barriers to organizational change, San Francisco: Jossey Bass.

[44] Bamber, C., Sharp, J., and Hides, M., (1999), "Factors Affecting Successful Implementation Of Total Productive Maintenance: A UK Manufacturing Case Study Perspective", Journal of Quality in Maintenance Engineering, Vol. 5, No. 3, pp. 162-81.

[45] Tsang, A.H.C., Chan, P.K., (2000), TPM implementation in China: a case study, International Journal of Quality & Reliability Management, 17, 2, 144-57.

[46] Santos, F., C., A., (2000), Integration of human resource management and competitive priorities of manufacturing strategy, International Journal of Operations & Production Management, 20(5), pp. 610- 628.

[47] Dangayach, G.S., Deshmukh, S.G., (2001), Manufacturing strategy: literature review and some issues, International Journal of Operations & Production Management, 21(7), pp. 884-932.

[48] Rudberg, M., 2002, Manufacturing strategy: linking competitive priorities, decision categories and manufacturing networks, PROFIL 17, Linkoping Institute of Technology, Linkoping.

[49] Schroeder, D.M., Congden, S.W., Gopinath, C., (1995), Linking competitive strategy and manufacturing process technology, Journal of Management Studies, 32(2), pp. 163-189.

[50] Nakajima, S., (1988), Introduction to Total Productive Maintenance, Productivity Press, Cambridge, MA.

[51] Hipkin, I., (2001), Knowledge and IS implementation: case studies in physical asset management, International Journal of Operations & Production Management, 21(10), pp. 1358-1380.

[52] Johansen, J., and Riis, J. O., (2005), The interactive firm - towards a new paradigm: A framework for the strategic positioning of the industrial company of the future, International Journal of Operations & Production Management, 25(2), pp. 202-216.

[53] Ramamurthy, K., (1995), Moderating influences of organizational attitude and compatibility on implementation success from computer-integrated manufacturing technology, International Journal of Production Research, 32(10), pp. 2251-2273.

[54] Marquez, A. C., and Herguedas, A. S., (2004), Learning about failure root causes through maintenance records analysis, Journal of Quality in Maintenance Engineering, 10(4), pp. 254-262.

[55] Kuivanen, R., 1996, Disturbance control in flexible manufacturing, The International Journal of Human Factors in Manufacturing, 6(1), pp.41-56.

[56] El-Haram, M., (1995), Integration approach to condition-based reliability assessment and maintenance planning, Ph D. Thesis, University of Exeter.

[57] Jarvinen, J., Vannas, V., Mattila, M., Karwowski, W., (1996), Causes and safety effects of production disturbances in FMS installations: a comparison of field survey studies in the USA and Finland, The International Journal of Human Factors in Manufacturing, 6, pp.57-72.

[58] Willmott, P.,(1997), Total Productive Maintenance: the Western Way, Butterworth-Heinemann, Oxford.

Acknowledgments

This paper is a part of the research conducted for CRC Integrated engineering Assets Management (CIEAM), at the University of South Australia. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the CIEAM.

CONDITION MONITORING OF BRIDGES BY VIBRATION ANALYSIS USING FIBER OPTIC INCLINOMETERS

Y Ishikawa[a] S Furukawa[a] E Sasaki[b] S Miyazaki[a] & H Yamada[b]

[a] *First Public Administration Systems Sector, NTT DATA Corporation, Toyosu Center Bldg., 3-3-3, Toyosu, Koto-ku, Tokyo 135-6033, Japan.*

[b] *Faculty of Environment and Information Science, Yokohama National University, 79-7 Tokiwadai, Hodogaya, Yokohama 240-8501, Japan.*

Abstract: We conducted experiments on a bridge to evaluate the ability of fiber optic inclinometers to detect changes in two kinds of vibration characteristics: the characteristic frequencies of steady-state vibration and the average maximum amplitude of impulsive vibrations caused by the passage of heavy vehicles. The experiments were performed on an I-beam girder bridge with tuned mass dampers. The dampers were turned on and off to change the target characteristics. The results showed that the changes in the target characteristics could be detected from the vibration measurements of the inclinometers. Fiber optic inclinometers are thus an effective tool for monitoring bridge health, not only in disaster situations but also under normal conditions.

Key Words: FBG inclinometers, health monitoring, fiber optic sensors, vibration analysis, bridge

1 INTRODUCTION

In a disaster situation, such as an earthquake or flood, many bridges spread out over a wide area may suffer various levels of damage. Safe evacuation of residents and timely delivery of relief supplies requires accurate and quick assessment of the structural damage to each bridge. Under normal conditions, bridges suffer continuous damage from the passage of heavy vehicles and the deterioration of their structural materials over time, so regular maintenance is needed. To meet the information needs for both situations, there is an increasing demand for real-time information gathering, continuous remote surveillance, and quantitative assessment. That is, there is a strong need for bridge health monitoring using sensors and networks[1]. Fiber optic sensors are well suited for health monitoring systems[2,3] because they reduce the number of transmission cables and improve the reliability of data transmission systems compared with conventional mechanical and electrical sensors. This is because many fiber optic sensors can be connected in series through a single optical fiber in 20km without a relay and can be used even without a power supply. Remote surveillance of the girder inclinations of bridges through optical fiber networks enables information about structural damage to be quickly gathered when a disaster strikes, making it possible to judge whether it is safe for vehicles to pass over the monitored bridges.

Fiber Bragg Grating (FBG) sensors, one of the most exploited fiber optic technologies, have excellent sensitivity and frequency characteristics, making them well suited for bridge health monitoring through vibration measurement[4,5]. The measured vibration can be used to identify internal damage at a very early stage. For example, a decrease in the characteristic frequency of a beam means a loss of beam stiffness. Using FBG inclinometers to measure the inclination and vibration of beams enables bridge conditions to be monitored using single-fiber optic sensors in both disaster and normal situations. This sensor system simplification reduces the cost of bridge monitoring systems.

We have evaluated the effectiveness of FBG inclinometers for monitoring bridge health. We experimentally measured the changes in the average maximum amplitude and the characteristic frequencies when the tuned mass dampers on a bridge were turned on and off. Comparison of the estimated characteristics with those obtained from simultaneous tri-axial acceleration measurements showed that vibration characteristics in a particular horizontal direction can be measured by matching the measurement direction of the FBG inclinometer to the target direction.

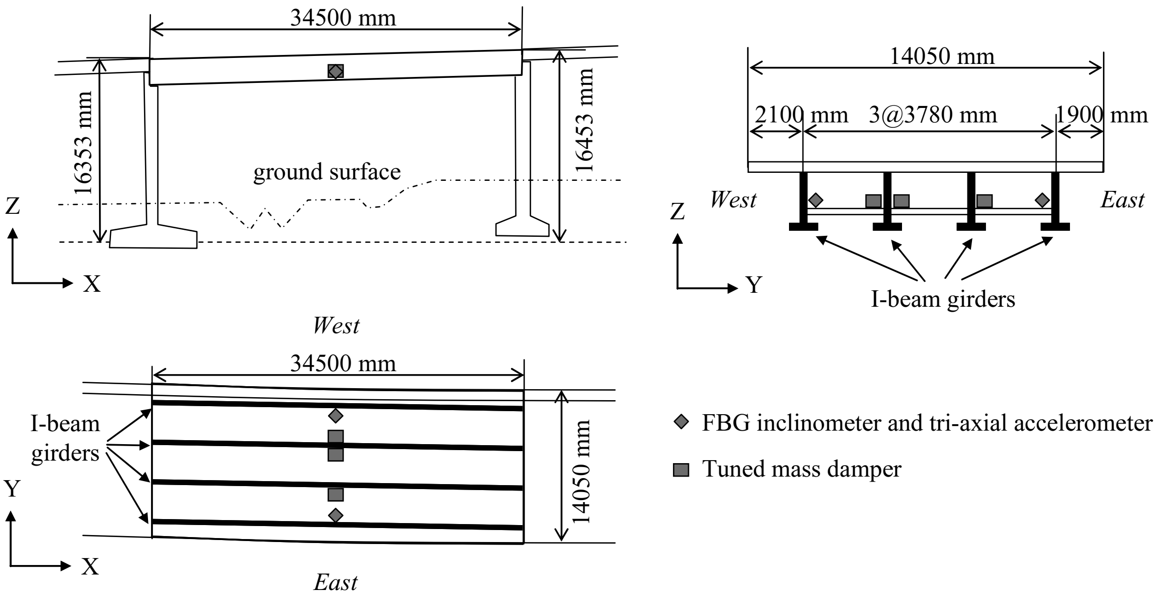


Figure 1. Arrangement of sensors and tuned mass dampers on target bridge.

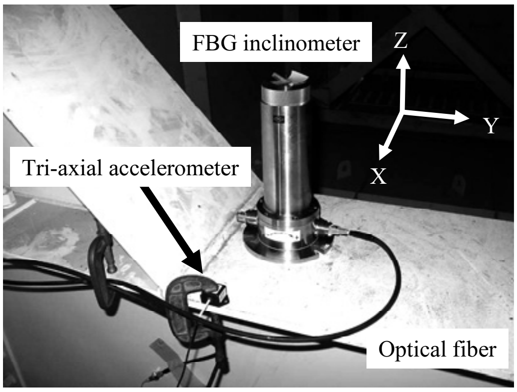


Figure 2. Sensor installation on east side.

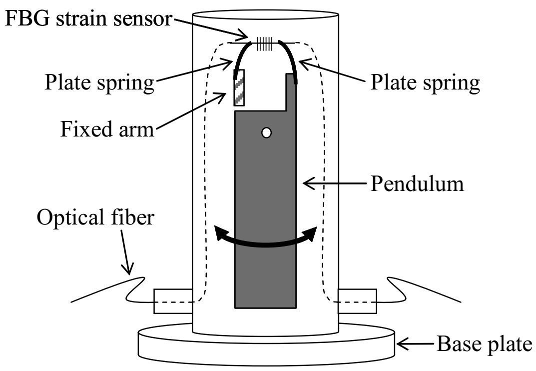


Figure 3. FBG inclinometer mechanism.

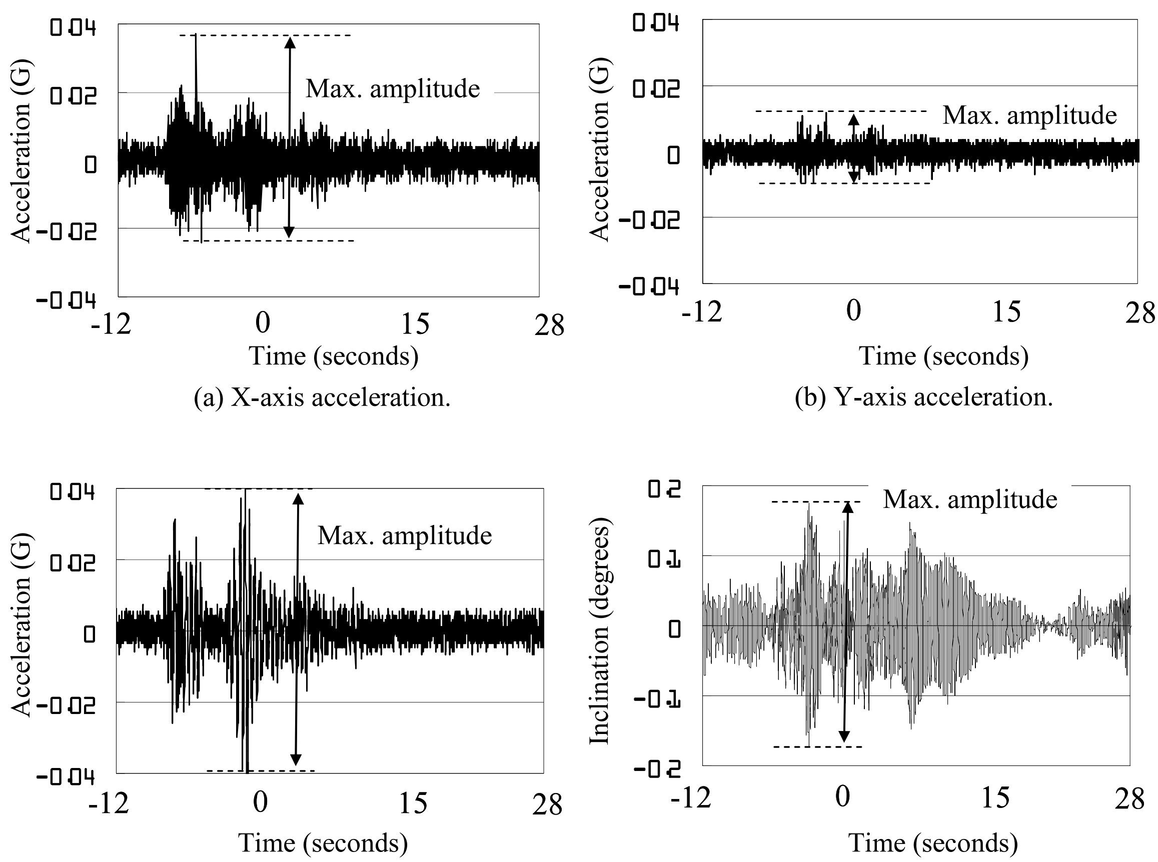
2 VIBRATION MEASUREMENT SCHEME AND SENSOR SETUP

We focused on two kinds of vibration characteristics: the average maximum amplitude of impulsive vibrations caused by the passage of heavy vehicles and the characteristic frequencies of the beams. An increase in the maximum amplitude indicates a decline in shoe functionality. Over the long term, however, the average of the maximum amplitude may increase due to an increase in the number of heavy vehicles crossing the bridge. Since such an increase would be a cause for concern, an increase in the average amplitude would indicate the need for an inspection. On the other hand a decrease in the characteristic frequency of a beam indicates a decrease in beam stiffness due to cracks, material deterioration, and so forth.

One way to detect damage due to an incident (e.g. a crack initiation) is to compare the vibration characteristics obtained for the bridge before and after the incident. This entails conducting periodic examinations, often over several years, until an incident occurs. To overcome this problem, we conducted our experiments using an I-beam girder bridge on which vibration-damping devices (tuned mass dampers, TMDs) had been installed (see Figure 1). The vibration characteristics measured when the TMDs were turned on were taken as pre-incident data, and those measured when the TMDs were turned off were taken as post-incident data. In both cases, the measurements were made using a single FBG inclinometer set on a cross beam connecting the I-beams. The ability of the FBG inclinometer to detect changes in the vibration characteristics was estimated from the measured inclination data. To verify the results, we set a tri-axial accelerometer next to the FBG inclinometer (see Figure 2).

The bridge is a part of a three-lane expressway and is crossed by about 60,000 vehicles every weekday. It has three TMDs on its cross beams at the centre of the span to reduce environmental vibration. As shown in Figure 1, an FBG inclinometer and a tri-axial accelerometer were installed on both sides of the span at the centre, the points of the bridge that suffer the most vibration due to deflection and torsion of the girders. These sensor positions are referred to as 'east side' and 'west side', and the direction of vibration to be measured is indicated by the X-, Y-, or Z-axis (see Figure 2). The sampling rate for the inclinometers was 250 Hz, and that for the accelerometer was 180 Hz. These rates are the highest available for these types of sensors.

As shown in Figure 3, the FBG inclinometers have a mechanism for generating strain in proportion to the inclination. It consists of a pendulum and a plate spring. The strain is measured with an FBG strain sensor. We call the direction in which the pendulum swings the 'measurement direction'. With this mechanism, the vibration of the body of the FBG inclinometer can be measured with the strain sensor even if the body is vibrated horizontally or vertically without being inclined. Here, however, data measured with the inclinometers is called the 'inclination'. Only one inclinometer and accelerometer pair was used in each measurement experiment. There were eight measurement patterns, which were the various possible combinations of the three measurement conditions: the sensor pair position (east or west), the inclinometer measurement direction (X- or Y-axis), and the TMD status (on or off). Each measurement was conducted for 24 hours on a weekday to make the traffic volumes comparable. The measurements thus took a total of eight days.



(a) X-axis acceleration. (b) Y-axis acceleration.

(c) Z-axis acceleration. (d) X-axis inclination.

Figure 4. Vibration data measured with tri-axial accelerometer and FBG inclinometer.

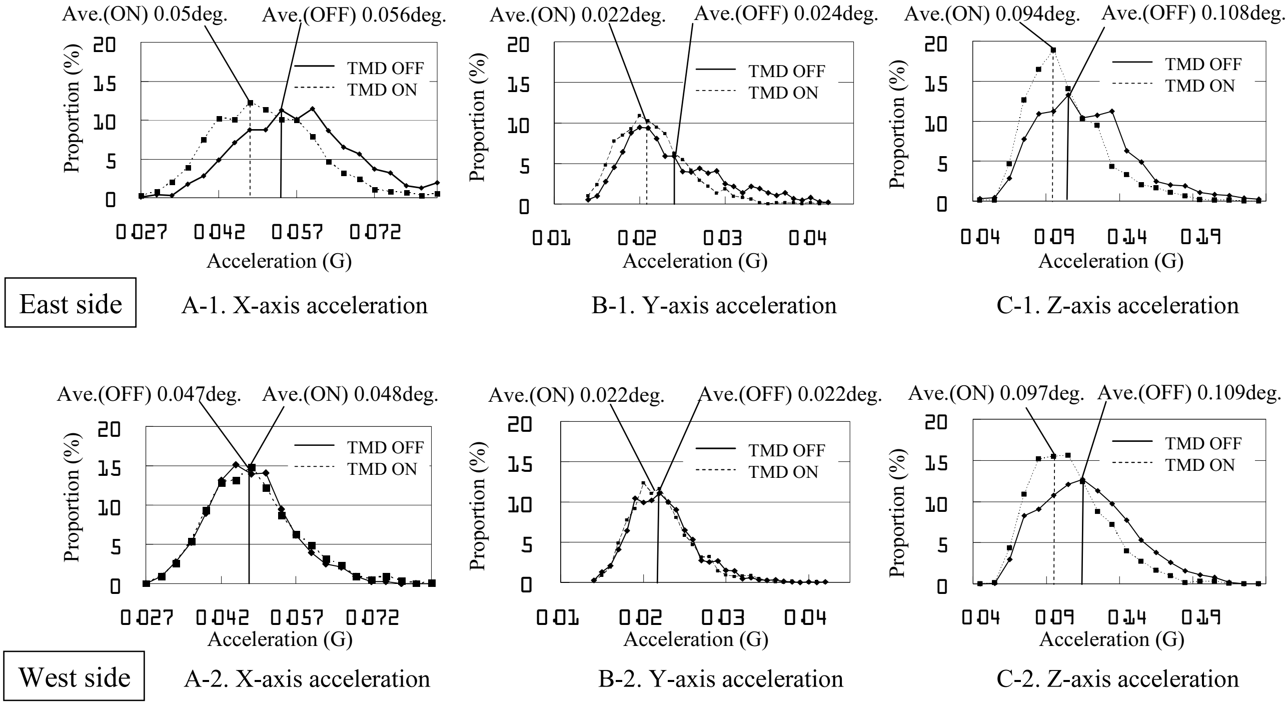


Figure 5. Histograms of maximum acceleration amplitude.

3 DETECTION OF CHANGES IN VIBRATION CHARACTERISTICS

3.1 Average maximum amplitude

To obtain vibration data when vehicles were crossing the bridge, we repeated the following two steps until all Z-axis acceleration data were scanned: (1) a time T at which the acceleration was either over 0.04 G or under -0.04 G was found in the Z-axis acceleration and (2) the data recorded from T−12 to T+28 seconds was extracted from the acceleration and inclination data. This set of data is shown in Figure 4. The total of data sets for the eight measurement patterns was 52,297 sets, so the average number per pattern was 6537.1, which is about 10% of that for the average traffic volume. Considering that a heavier vehicle creates a stronger effect on a girder, these data sets represent the passage of the 10% heaviest vehicles.

The maximum amplitude was calculated as the difference between the maximum and minimum values as shown in Figure 4. Ideally, the maximum amplitudes should be compared for data gathered when the same vehicle crosses the bridge in the different measurement patterns. This is problematic, however, because no other vehicles should be on the bridge at the same time as the measurement vehicle, and the bridge is subject to heavy traffic. Since the distribution of acceleration is usually approximately constant for the bridge, we compared the average maximum amplitudes in the measurement patterns, focusing particularly on the differences between the TMD-on and -off patterns.

Figure 5 shows the histograms and averages of the maximum acceleration vibration amplitudes. ‘TMD OFF’ means that the TMDs were off, and ‘TMD ON’ means they were on. As shown in A-1, B-1, and C-1, turning the TMDs on reduced the average for the east side by over 10% for each axis. Comparing the TMD-OFF and TMD-ON histograms in C-1 shows that above 0.11 G the average decreased, while there were marked increases at around 0.09 G due to turning on the TMDs. For the west side, the average maximum amplitudes for the Y- and Z-axes (B-2 and C-2) decreased while the averages for the X-axis increased slightly (see A-2)

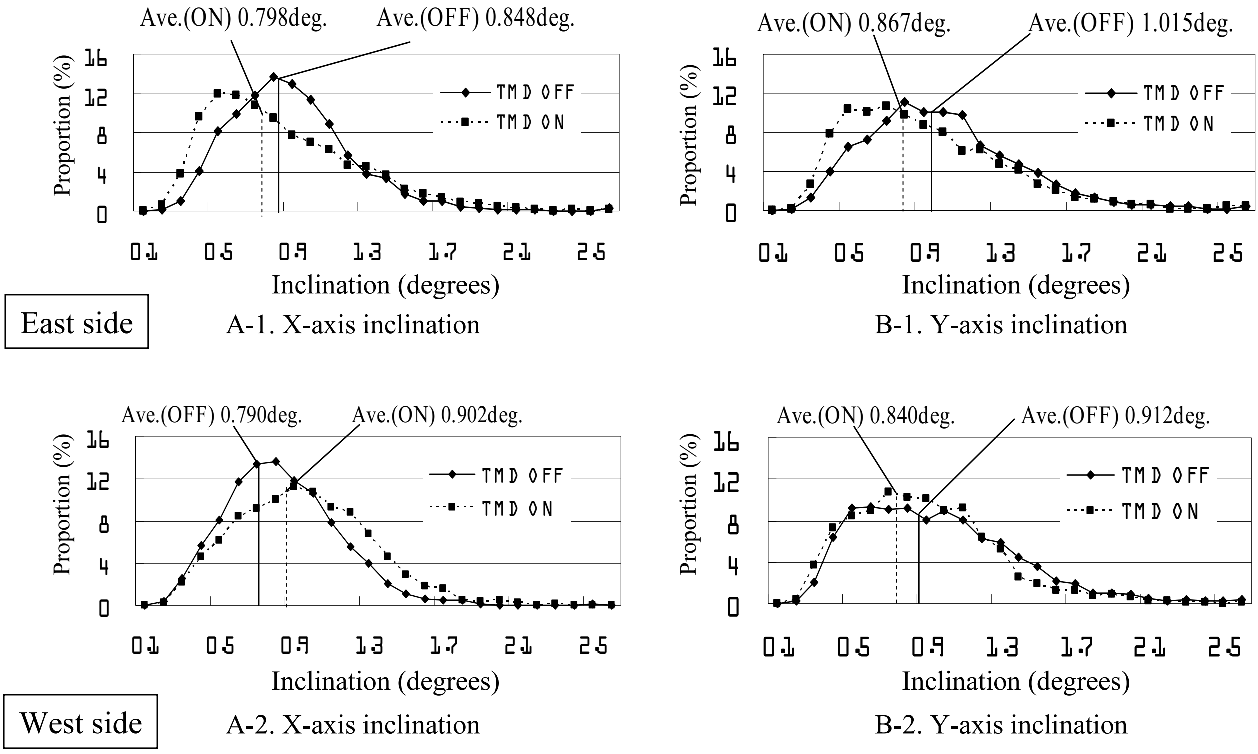


Figure 6. Histograms of maximum inclination amplitude.

Table 1: Decreases in average maximum amplitudes when TMDs were on.

		Acceleration (G)			Inclination (degrees)	
		X	Y	Z	X	Y
East side	OFF − ON	0.00639	0.00262	0.0136	0.05	0.148
	(OFF − ON) / OFF × 100	11.32%	10.83%	12.65%	5.9%	14.58%
West side	OFF − ON	−0.000625	0.000294	0.0119	−0.112	0.072
	(OFF − ON) / OFF × 100	−1.33%	1.31%	11.01%	−14.18%	7.89%

Figure 6 shows the histograms and averages of the maximum inclination vibration amplitudes. (The meanings of ‘TMD ON’ and ‘TMD OFF’ are the same as in Figure 5.) Comparing the TMD-OFF and TMD-ON histograms shows that the average maximum amplitude decreased in every case except when the FBG inclinometer was set on the west side in the X-axis direction (A-2).

The amount and ratio of decreases in the average maximum amplitudes when the TMDs were on are summarized in Table 1 for acceleration and inclination. For both, only the X-axis vibration on the west side increased when the TMDs were turned on. The tendencies of the changes in the X-axis and Y-axis averages thus coincided. This coincidence means that an FBG inclinometer can reliably detect change in the maximum amplitude of horizontal vibration. However, there is no clear quantitative relation between the decreases in acceleration and inclination.

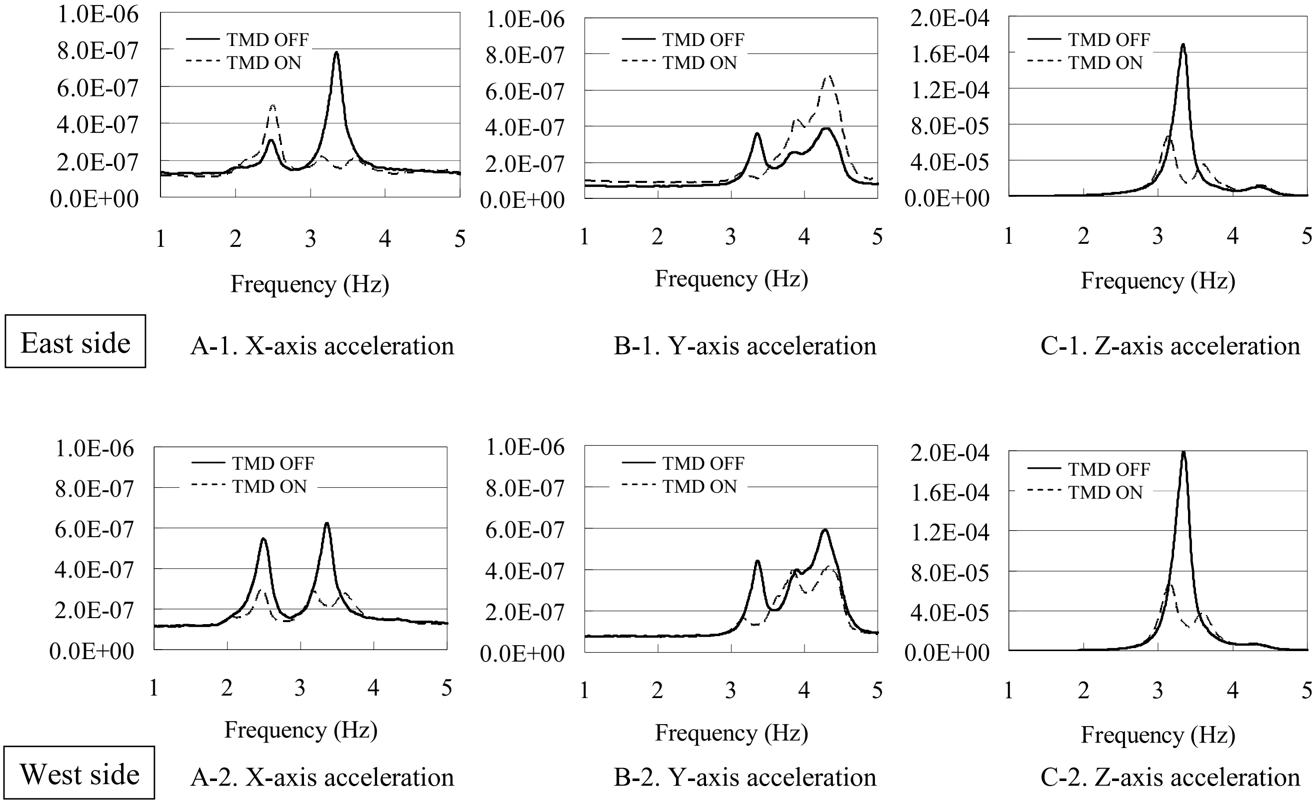


Figure 7. Power spectral densities of tri-axial acceleration.

3.2 Characteristic frequencies

To obtain the characteristic frequencies, we calculated the power spectral density for the acceleration data and the inclination data using fast Fourier transformation (FFT) and then identified peaks in the spectra. Preliminary testing had shown that identifying the peaks in the spectra for inclination vibration is difficult when the vibration amplitude is large. This is because the power of the characteristic vibrations was much smaller than that of the inclinometer pendulum, so that the peaks related to the characteristic vibrations of the bridge were not clear in the spectra. We thus used the steady-state vibration for the inclinometer measurement data. We did this by extracting the steady-state vibration portions of the inclination data. Each portion consisted of 4096 inclination values, each one less that 0.04°, for each of the eight measurement patterns. The power spectral density was then obtained by FFT for each portion. Finally, the power spectral density for each pattern was acquired as the average of the power spectral densities of the inclination parts measured for each pattern.

Figure 7 shows the power spectral densities of the tri-axial acceleration. Note that the vertical axes in C-1 and C-2 range to 2×10^{-4}, 200 times higher than the maximum in the other graphs, because the vertical acceleration was much stronger than the horizontal acceleration. A few characteristic frequencies can be perceived as clear peaks in these graphs. And the following features are found in the characteristic frequencies:

(a) For every case when the TMDs were off, there is a peak at 3.4 Hz. This peak disappeared when the TMDs were on.

(b) For the X-axis vibration, the bridge had a characteristic frequency at 2.4 Hz besides 3.4 Hz.

(c) For the Y-axis vibration, the bridge had characteristic frequencies at 3.9 and 4.4 Hz besides 3.4 Hz.

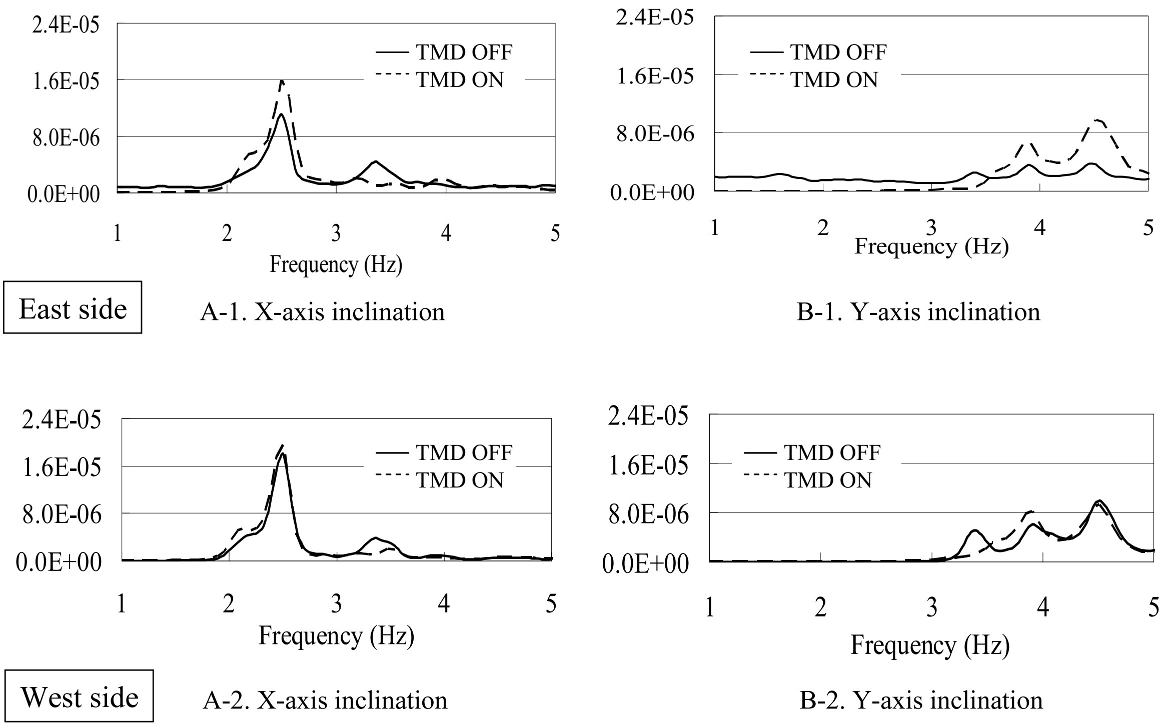


Figure 8. Power spectral densities of inclination.

The same findings hold for the power spectral densities of the inclination, shown in Figure 8. The first one means that an FBG inclinometer can detect changes in the characteristic vibration. The second two mean that an FBG inclinometer can acquire the characteristic frequencies of vibration in either the X- or Y-axis direction if its measurement direction is adjusted accordingly. As mentioned above, the vertical vibration (Z-axis vibration) at 3.4 Hz was about 200 times as strong as the horizontal vibration when the TMDs were off. The horizontal vibration also included characteristic vibration at 3.4 Hz. However, in the power spectral densities of inclination, the vibration at 3.4 Hz was weaker than the other characteristic frequencies at 2.4, 3.9, and 4.5 Hz. This means that it is possible to obtain the characteristic frequencies of horizontal vibration by using an FBG inclinometer without interference from vertical vibration.

4 DISCUSSION

Our comparison of tri-axial acceleration and inclination with regard to the vibration characteristics (average maximum amplitude and characteristic frequency) showed that it is possible for an FBG inclinometer to detect changes in the vibration characteristics caused by turning the TMDs on and off. This means that FBG inclinometers can be used to detect damage to bridges if the damage caused an increase in the vibration amplitude or a decrease in the characteristic frequency.

The experimental results for the average maximum amplitude and characteristic frequency make it clear that measurement using FBG inclinometers can reveal the vibration characteristics in the corresponding measurement direction. Even if only one FBG inclinometer is available, it is easy to change its measurement direction because the installation is on a level surface, unlike FBG strain sensors, which are stuck on surfaces with adhesive. If more than one FBG inclinometer is available, the vibration characteristics in multiple directions can be monitored at the same time because the FBG sensors can be connected in series through a single optical fiber.

However, the measurement directions of FBG inclinometers are obviously limited to horizontal ones. The ability of FBG inclinometers to detect the characteristics of vertical vibration could not be determined from our experimental results. Considering that vertical vibration is stronger than horizontal vibration, as shown in Figure 7, we need to develop a method of using FBG inclinometers to detect the characteristics of vertical vibration. In addition, inclined vibration should also be considered because inclination directly affects the measurement results. Comparison between the acceleration and inclination in the latter half of the measurement data shown in Figure 4 shows that the inclination amplitude changed while the acceleration ones were steady. The acceleration in our experiments was mainly caused by the passage of vehicles on either the east or west side of the bridge, the locations of the accelerometers. However, changes in inclination may be caused by the passage of vehicles on both the east and west sides because inclination is affected by torsion, i.e. rotation on the X-axis. This difference between the measurement characteristics of FBG inclinometers and tri-axial accelerometers may account for the difference shown in Figure 4. Therefore the measurement characteristics of FBG inclinometers for vertical and inclined vibration need to be examined.

5 CONCLUSIONS

We have demonstrated that FBG inclinometers can be used to detect changes in the vibration characteristics of bridges and that FBG inclinometers are an effective tool for monitoring bridge health. Experimental measurements in which two kinds of vibration characteristics (average maximal amplitude and characteristic frequency) were estimated with and without the effect of tuned mass dampers showed that changes in these characteristics are detectable using FBG inclinometers. Comparison of the estimated characteristics with those obtained from simultaneous tri-axial acceleration measurements showed that vibration characteristics in a particular horizontal direction can be estimated by matching the measurement direction of the FBG inclinometer to the target direction.

Future work includes using FBG inclinometers to estimate the characteristics of perpendicular and inclined vibrations. To clarify the measurement characteristics of FBG inclinometers for both kinds of vibrations we plan to conduct vibration tests.

6 REFERENCES

[1] Ko JM & Ni YQ. (2005) Technology developments in structural health monitoring of large-scale bridges. Engineering Structures, 27(12), pp. 1715–1725.

[2] Udd E. (2002) Good Sense. SPIE OE Magazine, 2(8), pp. 27–30.

[3] Ishikawa Y, Miyazaki S, Uehara H & Inoue U. (2004) Design of Integrated Heterogeneous Sensor Systems for Disaster and Environmental Monitoring. In Proceedings of the 1st International Workshop on Networked Sensing Systems, Tokyo. pp. 113–116.

[4] Xu B & Wu ZS. (2003) Long-gage fiber optic sensors for dynamic strain measurement and structural identification. In Wu & Abe. (Eds) Proceedings of the Structural Health Monitoring and Intelligent Infrastructure, Tokyo. pp. 299–307.

[5] Macdonald JHG & Daniell WE. (2005) Variation of modal parameters of a cable-stayed bridge identified from ambient vibration measurements and FE modelling. Engineering Structures, 27(13), pp. 1916–1930.

7 ACKNOWLEDGMENTS

We are grateful to the members of the Sensor Network Group of NTT DATA Corporation for supporting this work.

REMOTE MACHINE MONITORING THROUGH MOBILE PHONE, SMARTPHONE OR PDA

W Wanbin and PW Tse

Smart Engineering Asset Management Laboratory, Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Tat Chee Ave., Hong Kong, P. R, China

Abstract: Machine condition monitoring is important to the efficiency of a factory and the safety of its workers. A variety of vibration signals analysis techniques have long been using to diagnose machine's status. Since the Web has been invented, researchers have begun to explore web-based machine diagnostic system. Today, the GSM (Global System for Mobile Communication) and the CDMA (Code Division Multiple Access) have already been widely deployed. Mobile communication provides a variety of data services. Java program can be run in any Smartphone or PDA (personal digital assistant). In this paper, we introduce an innovative application of machine fault diagnosis using mobile communication. For example, our system can send warning acknowledgements via short messages. It can also automatically call the mobile phones of concerned engineers to alert them the abnormal status of the monitored machines. Through our system, maintenance managers can obtain machine curtailed data through their mobile phones using WML (Wireless Markup Language) so that they can make managerial decisions based on the machine condition instantaneously. For engineers, they can obtain detailed data through java program running in a Smartphone or PDA so that they can perform machine fault diagnosis correspondingly.

Key Words: Machine fault diagnosis, Remote monitoring, Mobile communication, Web-based applications

1 INTRODUCTION

Machine condition monitoring has been a popular research topic since machinery has been invented. Defects that occur in a machine always exhibit a symptom in the form of vibration or other parameters. There are machine maintenance techniques such as: vibration monitoring, oil analysis, particle analysis, thermography etc. Vibration monitoring is the most effective technique to detect mechanical defects in rotating machinery. When a machine is starting to deteriorate, the forces acting on different components of the machine will be changed. The corresponding vibration spectrum of the machine will also change. Hence, it is possible to using signals analysis techniques to detect the occurrence of faults in the machine. A variety of vibration signals analysis techniques have long been using to diagnosis machine health status through analysing its generated vibration spectrum. These techniques include Fourier Transform, Wavelet transform, neural networks etc. Numerous researches have been conducted in the field of machine fault diagnosis.

On the other hand, the trend of manufacturing is moving toward regionalization. The production machines, which need to be monitored, could be located at a plant that is far away from the office of the managers and fault diagnosis experts. Since the day that the Web has become a stable avenue for idea exchange, researchers began to explore web-based machine fault diagnostic system. Caldwell et al. (1998) described a system, which used the Internet for diagnosing a scanning electron microscope. Ong et al. (2001) presented a Web-based knowledge-based fault diagnostic and learning system which used multi-agent technology. Yuan et al. (2004) described the architecture of a Web-based remote operation and monitoring system served for open NC (Network computer) devices. Wang et al. (2004) employed virtual instruments to provide remote sensing, monitoring and on-line fault diagnosis for equipment, together with a collaborative maintenance platform for international experts to interactively share their experiences in maintenance.

Nowadays, mobile communication is very popular. Messages, voices, even video clips can be sent remotely via mobile communication and received by a mobile phone or PDA. The GSM (Global System for Mobile Communication) and CDMA (Code Division Multiple Access) have already been widely deployed in mobile communication. Wireless Application Protocol (WAP) has also been developed for mobile devices, such as mobile phones and PDAs, to access the Internet. On the other hand, the developments of Smartphone and PDA technologies allow the transformation of a mobile terminal into a mini computer, and make it possible to process data within a mobile phone/PDA. Sun Microsystems provide free binary implementations to their Java ME runtime environment which is supported by all types of Smartphones and PDAs. Hence, the mobile phone grows from a voice communications device into a multipurpose information terminal. The development of these technologies makes it possible to maintain machines from anywhere and at anytime.

In this paper, we introduce our development on the application of mobile communication for machine fault diagnosis. The paper is laid out in four sections with this section as the introduction. Section 2 gives some research background about the development of mobile communication. Section 3 presents the architecture and layout of our mobile communication based machine fault diagnostic system. It describes in details about the automatic warning acknowledgement system. The system can automatically send alert messages to the maintenance managers/engineers' mobile phones, and also call their phones to make sure the managers/engineers have acknowledged the alert messages when the health status of a monitored machine becomes abnormal. The system will publish the information of the machine status in a Web/WAP server. Then, managers get machine curtailed data through WML by mobile phone; engineers get diagnosed data through java program running in the Smartphone or PDA. Finally, section 4 provides the concluding remarks and the future development.

2 THE DEVELOPMENT OF MOBILE COMMUNICATION

In1979, the first commercial mobile phone network was opened for business in Tokyo. This is a cellular telephone system. By the mid-1980s there was a significant expansion of mobile services offered to the general public in the world. Customers were able to avail themselves of the service using a mobile telephone the size of a brick. There are several different mobile systems such as NMT, AMPS, TACS, RTMI, C-Netz, and Radiocom 2000. These systems later became known as first generation (1G) mobile phones.

In the 1990s, second generation (2G) mobile phone systems such as GSM, IS-136 ("TDMA"), iDEN and IS-95 ("CDMA") began to be introduced. Phase I of the GSM (global system for mobile telecommunications) specifications were published in 1990 and the first GSM network was officially opened in Finland on July 1, 1991. 2G phone systems were characterised by digital circuit switched transmission and the introduction of advanced and fast phone to network signalling. GSM provides a variety of data services. Users can send and receive data, at rates up to 9600 bps. A unique feature of GSM, not found in older analog systems, is the Short Message Service (SMS). SMS is a bidirectional service for short alphanumeric (up to 160 bytes) messages.

Not long after the introduction of 2G networks, projects began to develop 3G systems. Inevitably there were many different standards with different contenders pushing their own technologies. Quite differently from 2G systems, however, the meaning of 3G has been standardised in the IMT-2000 standardisation process. This process did not standardise on a technology, but rather on a set of requirements (2 Mbit/s maximum data rate indoors, 384 kbit/s outdoors, for example).

During the development of 3G systems, 2.5G systems such as CDMA2000 1x and GPRS were developed as extensions to existing 2G networks. These provide some of the features of 3G without fulfilling the promised high data rates or full range of multimedia services. For example, CDMA2000-1X delivers theoretical maximum data speeds of up to 307 kbit/s.

The Internet is a simple and efficient method of delivering services to millions of PC users. There is a requirement surfing Internet by mobile phone. Then WAP was specially developed to make the convenience of the Internet available to mobile users. The first commercial WAP services were based on the WAP 1.1 standard, issued in June 1999. WAP2.0, which was released in Summer, 2001, incorporates TCP/IP and xHTML into the WAP standard, offering programmers the facility to develop an application fro both fixed and wireless Internet at once.

The Micro Edition of the Java 2 Platform provides an application environment that specifically addresses the needs of commodities in the vast and rapidly growing consumer and embedded space, including mobile phones, personal digital assistants etc.

Today, 2G or 2.5G mobile systems have already been widely deployed. In some country, users even can choose 3G services. The development of these technologies gives a new research topic of machine fault diagnosis system.

3 THE APPLICATION OF MOBILE TECHNOLOGY IN FAULT DIAGNOSIS

The application of mobile technologies for our machine fault diagnostic system is showed in Figure 1. SAMS (Smart Asset Maintenance Systems) uses LabVIEW development tools for programming and provides a graphical user interface (GUI) that allows users to choose different virtual instruments for machine fault disgnosis (Leung & Tse 2005). SAMS has a number of signal processing techniques for machine fault diagnosis, such as the Fast Fourier Transform (FFT), Continuous Wavelet Transform (CWT), Exact Wavelet Analysis (EXA) (Tse et al. 2004), Orbit Analysis etc. The hardware part of SAMS includes sensors, Web-camera or video-camera, microphones, signal amplifiers or conditioners, data acquisition (DAQ) cards. The sensors collect the running signals from the monitored machine and then transmit the collected signals to the amplifiers. The amplifiers condition the signals into acceptable formats so that the DAQ cards can convert them to digital signals for further fault analysis.

When SAMS detects the health status of the monitored machine becomes abnormal, the Java program running in the server will send short messages and call the concerned engineers' phones through the mobile connected to the Web/Wap server. SAMS will analyse the collected data, perform the necessary diagnoses, and then transmit the diagnostic results and data to the Web/Wap server. The Java program runs in the server will generate the WML file that contains diagnostic results. Users can get machine curtailed data by mobile phones. Engineers will also obtain the detailed data and diagnostic results through the java program running in their Smartphones or PDAs.

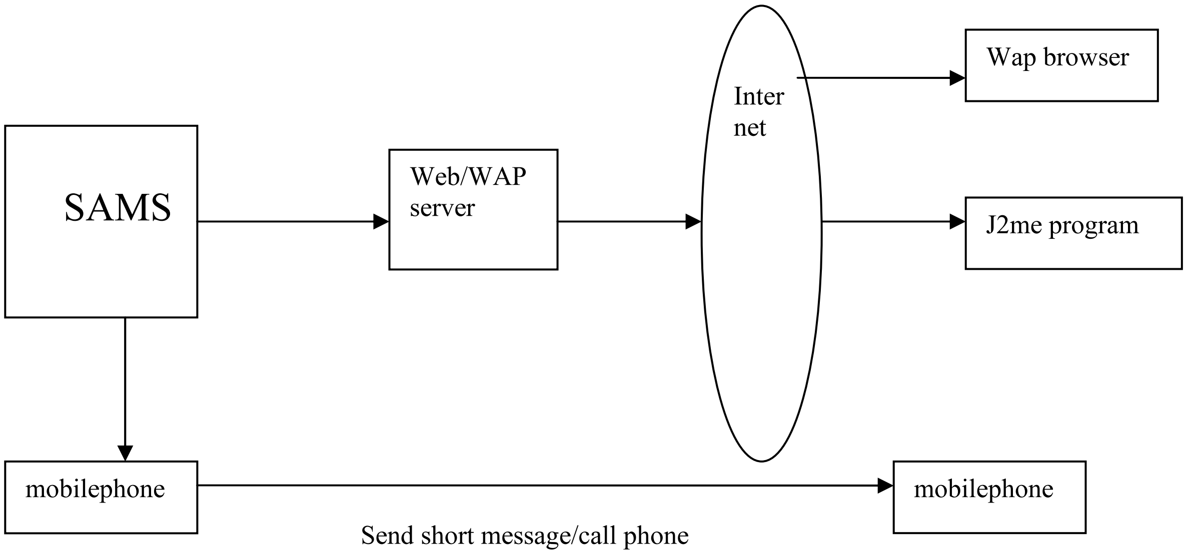


Figure 1. The architecture of mobile communication for our remote machine fault diagnostic system

3.1 Development of the automatic alarm acknowledgment

Short Message Service (SMS) is a service available on most digital mobile phones that permits the sending of short messages between mobile phones. SMS was originally designed as part of the GSM network, but is now available on a wider range of networks, including the 3G network. Using SMS in industry is not a new idea. Wu and Jan (2003) proposed a Home Network System integrated with WAP and SMS to support the connectivity between home and the Internet for mobile communication networks. Al-Ali et al. (2004) presented a system which allows the homeowner to monitor and control his house appliances via his mobile phone by sending commands in the form of SMS messages while receiving the appliances operating status. These systems use the AT commands to send SMS through a GSM modem. AT commands are easy to understand and implement. However, the functions of AT commands is limited.

Our research is based on the Microsoft's Smartphone 2003 smart-phone operating system. It is a revolutionary mobile phone software platform that allows a variety of functions rather than just a phone for talking and listening. It has empowered the developers to build rich and innovative mobile applications by providing powerful and flexible development tools. Developers can target Windows Mobile natively with Visual C++. They can also target the .NET Compact Framework with managed code by using C# or Visual Basic .NET.

Our system can detect an abnormal operation condition of the monitored machine and then issue a warning to users. If such a case occurred, the Java Program will run the programs in Motorola MPX 220 mobile phone that is connected to the server. There are two programs in the mobile phone. One program will send messages to the user mobile phone and the other program will call the user mobile phone. The two programs are coded by Visual C#. The first program needs the user mobile phone numbers and message contents. The second program only needs the mobile phone numbers.

Below is part of the java program called 'Callphone', which is used to call the user mobile phone.

```
cmdString1="rapistart \\Storage\\callphone"+"  "+CallPhonenum;

Process proc2 = Runtime.getRuntime().exec(cmdString1);
```

The command 'rapistart' is used for remotely starting an application on a PDA or Smartphone. It is a part of the Windows Mobile Developer Power Toys which can be downloaded from Microsoft's website. The command 'callphone' is an exe program which has been installed in the Motorola MPX 220. The command 'Storage' is a directory where the Callphone program is saved. Callphone is coded by Visual C# and uses Smartphone 2003 API to make a call. The command 'CallPhonenum' is the phone number which is retrieved from the text file saved in the computer. The command 'Runtime.getRuntime().exec()' is a java method that executes the specified string command. Here, it executes the command 'rapistart' to run the program in the MPX200 for calling a specified mobile phone.

The Java program will send short messages containing the current health status of the machine to engineers and managers. Figure 2 gives a sample of the message that the engineers and managers will receive when the monitored machine becomes abnormal. It includes the name of the company or the factory, the machine identification, the monitored machine component, the time and date that the message is produced, and the current health status of the machine.

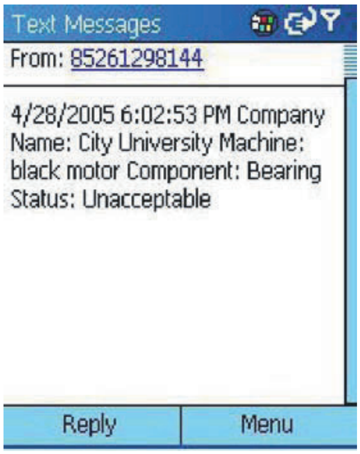


Figure 2. The SMS message sent to users to acknowledge them the abnormal status of the monitored machine.

3.2 Get machine health information via WAP browser

The Internet is a simple and efficient method of delivering services to millions of PC users. WAP was specially developed to make the convenience of the Internet available to mobile users. In 1997, several wireless-phone manufacturers organized an industry group to formalize the WAP standards. The standards are used to define communication protocol and implementation for wireless-network applications. WAP allows cell phones, PDAs, and other low-computational-power devices access to the Internet.

Services made up in the Hypertext Markup Language (HTML), which is used for the layout of Internet pages on the Web, do not take sufficient account of specific circumstances that arise in a mobile environment. The WML (Wireless Markup Language) is the mobile variation of HTML. It is designed for creating WAP based applications. It is user-interface independent. It supports text, images, user input, variables, navigation mechanisms, multiple languages, and management server requests. The WML has been designed to adapt high-latency and narrow-band types of wireless networks. When a user gains access to a WAP site, it will send contents in a deck of cards through which the user can browse. The WML is supported by almost every new mobile phone browser around the world.

Our system will write important machine health information via the WML and publish the information by a Web/WAP server. Figure 3 gives an example on how the important machine health status can be displayed in the small screen of a Motorola MPx-220 phone. It shows a menu for displaying the current health conditions of a machine in our laboratory. The cursor can be scrolled up or down to select a particular point to view its health status, and to get the real time image of the machine. The point's health data and the image of the monitored machine are available on the screen of the mobile phone/PDA.

The WML menu page, of which the user can obtain the machine's data, image, may seems very simple. However, many new technologies have been employed to implement such page. The system gets the collected vibration data and diagnostic results from the SAMS every 20 seconds (the interval can be reconfigurable as desired) and generates a WML file to contains the information. The WML file includes the menu page and the details of a selected sensor measuring point (referred to a particular component of the monitored machine). By choosing a particular sensor measuring point option, the remote user can get the detailed information of that measured point. The information includes the source of data, its values, the health status, and the reference used to determine its health status of the measured point or component. The health status can be classified as 'Good', 'Satisfactory', 'Unsatisfactory', and 'Unacceptable'. The reference (REF) in Figure 4 shows that the RMS (the root-mean-square value of the collected vibration data) is LT (less than) 0.2 mm/s (vibration in velocity unit). This means that the machine health status is good. If the RMS is greater than 0.8 mm/s, then the monitored machine's health may be classified as 'Unacceptable'. For demonstration purpose, we are only using the RMS value as reference. In future, we will use other machine health vibration indicators, such as kurtosis, crest factor, higher order statistical values, to synthetically determine the health of the monitored machine component.

On the menu page, by clicking the 'Machine Picture' link, the WAP browser will download the picture of the currently monitored component (as shown in Figure 5) from the Web/WAP server. The Java program running in the server uses JMStudio to capture an image of the machine every 30 seconds (the interval can be reconfigurable as desired) through a Web camera. JMStudio is a Java application that uses the JMF (Java Media Framework) 2.0 API to play, capture, transcode, and write media data. A VI program sends this JPEG file to the web server and then web server publishes it to the Internet.

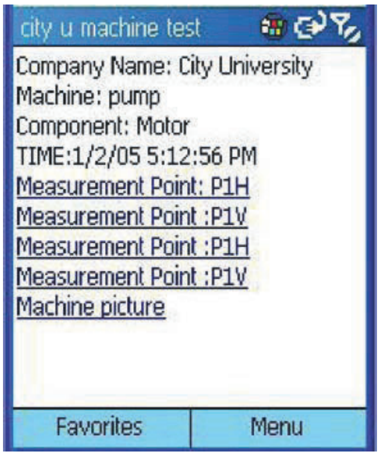


Figure 3. The content of the SMS that acquiring machine health status via mobile phone.

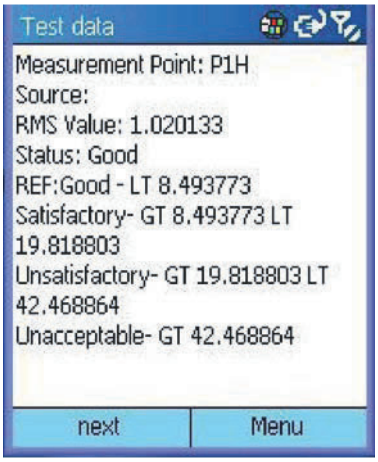


Figure 4. The health status of a particular machine component (through the choice of a sensor measuring point).

Figure 5. The image of the currently measuring point of the monitored machine.

3.3 Get machine data using Java program running in the Smartphone or PDA

Java is an object-oriented programming language that can be run on any platform. Java technology is used in applications on PCs, servers and is also applied to various embedded devices. Hoshi (1999) gave some examples of applications of Java in the manufacturing industry. In our research, we use java program which run in the Smartphone or PDA to get the machine data file. Figure 6 shows a demo for running the Java program in a device emulator of the J2ME Wireless Toolkit, which is a set of tool for creating Java applications that run on devices compliant with the Java Technology for the Wireless Industry. The URL is the default location of the machine data file. It can be changed by input different URL using the mobile phone keys. Figure 7 displays the content of the file that contains the collected machine vibration data. Because of the memory limitation of mobile devices, the machine data file cannot be too large.

Figure 6. The screen of a mobile phone showing the Web address to get the machine diagnostic information.

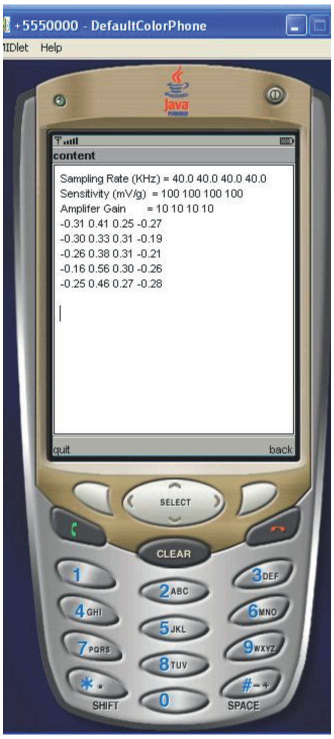


Figure 7. The curtailed vibration data collected from the monitored machine.

4 CONCLUSIONS

With the continuos improvements in machine condition based monitoring and diagnosis and the requirement of globalization in manufacturing, there is a strong demand on the use of remote machine health monitoring, especially through mobile communication, such as mobile phones and PDAs. In this paper, we introduce our development of using mobile communication to inform concerned users the health status of a machine which is located remotely from the users. The users can remotely and wirelessly obtain the current operating and health conditions of any manufacturing machine even without the access of a Web wired connection. Through our system, maintenance managers can obtain machine curtailed data through their mobile phones using WML (Wireless Markup Language) so that they can make managerial decisions based on the machine condition instantaneously. For engineers, they can obtain detailed data through java program running in a Smartphone or PDA so that they can perform machine fault diagnosis correspondingly.

Today, the developments in Smartphone and PDA allow the transformation of a mobile device into a mini computer, and make reasonably large amount of data to process within a mobile phone/PDA. Our future work will concentrate on the potential of analysing collected machine data directly in a Smartphone or PDA rather than from a Web/WAP server as we currently used. The third generation of mobile communication is expected to be capable of transmitting data of text, digitized voice, video, and other multimedia data at a rate of 2 megabits per second (Mbps) and higher. Hence, remote and wireless machine health monitoring and fault diagnosis through mobile communication will become a trend in modern maintenance and engineering asset management.

5 REFERENCES

[1] Al-Ali A.R, Al Rousan M., Mohandes M. (2004) GSM-based wireless home appliances monitoring and control system. Proceedings 2004 International Conference on Information and Communication Technologies: From Theory to Applications, ICTTA 2004, 237-238.

[2] Caldwell, Nicholas H., Breton, Bernard C., Holburn, David M. (1998) Remote instrument diagnosis on the Internet. IEEE Intelligent Systems & Their Applications, Vol. 13(3): 70-76.

[3] Leung J. and Tse P.(2005) Smart Asset Maintenance System for Machine Fault Diagnosis – Its Effectiveness, Methodology, and Applications, Proceedings of IDET/CIE 2005, ASME 2005 International Design Engineering Technical Conference, Symposium on Dynamic Signal Processing on Engineering System Monitoring and Diagnostics, California, USA, 24-28 Sept. 2005, Vol. 1, DETC2005-84300, CD-Rom, 8 pages.

[4] Hoshi, Tetsuo(1999) Current and future Java technology for manufacturing industry, Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, v 6, 1999,404-409.

[5] Ong K., An N., Nee A.(2001) Web-based fault diagnostic and learning system, International Journal of Advanced Manufacturing Technology, Vol.18(7):502-511.

[6] Tse P., Yang W.X., and Tam H. Y.(2004) Machine Fault Diagnosis Through an Effective Exact Wavelet Analysis, Journal of Sound and Vibration, Vol. 277(4-5), Nov. 5, 2004, 1005-1024.

[7] Wang J., Tse P., He L., Yeung R.(2004) Remote Sensing, Diagnosis, and Collaborative Maintenance with Web-enabled Virtual Instruments and Mini-Servers, International Journal of Advanced Manufacturing Technology, Vol. 24(9-10), Nov. 2004, 764-772.

[8] Wu Chi-Hsiang, Jan Rong-Hong,(2003) System integration of WAP and SMS for home network system. Computer Networks, v 42, n 4, Jul 15, 2003, 493-502.

[9] Yuan C., Zhang G., Chen Y., Zhou Z., (2004) Web-based remote operation and monitoring technology for open NC devices. Manufacturing Automation 2004-03,36-39.

6 ACKNOWLEDGMENTS

The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 1100/02E).

A SOPHISTICATED BUT EASY-TO-USE AND COST-EFFECTIVE MACHINE CONDITION MONITORING AND DEGRATION PREDICTION SYSTEM

Peter W.Tse [a] and Jacko T. Leung[b]

[a] *Director of Smart Engineering Asset Maintenance(SEAM) Laboratory, Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Tat Chee Ave., Hong Kong, P. R, China.,E-mail: meptse@cityu.edu.hk Tel: (852)27888431 Fax: (852)21942289*

[b] *Smart Engineering Asset Maintenance(SEAM) Laboratory, Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Tat Chee Ave., Hong Kong, P. R, China., E-mail: mejacko@cityu.edu.hk Tel: (852)21942651 Fax: (852)21942289*

Abstract: All kinds of machines require proper maintenance, especially when the operation of machines involves human beings. In order to minimize machines' failure rate, researchers have been searching more sophisticated methodologies for the earliest detection of the machine faults prior to causing any catastrophe. On the other hand, the industry requires such sophisticated methods easy-to-use and affordable. Hence, this paper presents a machine condition and its degradation monitoring system with the state-of-the-art capability yet easy-to-use and low-cost. In most of the rotating machines, bearings are the most frequent failure parts. Traditionally, maintenance persons would recognize the health of these parts through the monitoring of their vibration frequency spectra. However, modern machineries are so complex that many components may generate vibrations and affect each others. The vibration signals generated from bearing rolling parts may be overwhelmed by other higher amplitude vibration signals generated by nearby larger components. Hence, maintenance persons always encountered difficulties to distinguish the difference of vibrations generated by a bearing or nearby components. Hence, sophisticated machine fault diagnoses are required. Although a number of advanced diagnostic methods are available, they are either not ready for commercial use or too expensive and difficult to use without comprehensive learning. Therefore, we have designed and implemented an economic but yet efficient machine condition and its degradation monitoring system called the 'Smart Asset Maintenance System (SAMS)'. Here, we have described its effective functions on machine fault diagnosis and prognosis, added-valued capabilities, such as automatic report generation, smart data management, and ready to use Web-based remote monitoring. Most surprisingly, even equipped so many advanced functions, SAMS has been made for user-friendly and low-cost.

Keywords: Machine condition monitoring; vibration analysis; noise cancellation; vibration trend; remote sensing.

1 INTRODUCTION

All rotating machines generate noise and vibration, even small amplitudes of vibration can cause a large amount of damage in different parts of the machine. In ideal case, machine running as a prefect condition would not produce any vibration or sound it is because all the energy is transfer into the work to be done. However, in real life, vibration should occur when the machine is running and energy would excite through the structure in the form of vibration. As long as the excitation energy, the vibration would be a typical level and contain a characteristic frequency shape when the machine is in normal. And the analysis of the produced vibration can be used to give information on the condition of the machines. Thus it is important to monitor or record the vibration signal in order to distinguish the health condition. In fact, when the machine begin to be defective, the force acting on different components will be changed, which will result on changing of the vibration spectrum. Therefore, a variety of signal processing techniques have been applied in the machine monitoring to detect the change or defect of the machine. The conventional techniques include the Fast Fourier Transform (FFT), Envelop Detection, and Impact Test. Recently, a number of new signal processing techniques have been developed to enhance the accuracy and efficiency in machine fault diagnosis. Some of them are Continuous Wavelet Transform (CWT), Exact Wavelet Analysis (EWA), Active Noise Cancellation (ANC), Orbit Analysis, and Vibration Trend Prediction. In this essay, an innovative system for maintenance would introduce. The novel system is called the Smart Asset Maintenance System (SAMS), especially made for the industry to uplift the technology used in machine condition monitoring and fault diagnosis. SAMS has embedded with all the aforementioned technologies. In section 2, we have discussed the methodologies, both conventional and new ones that are embedded in SAMS. Besides, in section 3, the configuration of Data acquisition system was also introduced. At last, some of selected industrial applications using SAMS to monitor the machine health are also presented.

2 SELECTED DESCIPTIONS ON WEB-ENABLED VIRTUAL INSTRUMENTS OF SAMS

Vibration signals are usually consist of very many frequencies occurring simultaneously so that we cannot immediately see just by looking at the amplitude – time pattern (time waveform). Therefore, we need some signal analysis technology, such as the FFT and wavelet, so that we are able to track down the source of undesirable vibration.

2.1 Conventional maintenance algorithm

In tradition, FFT algorithm is used to calculate a spectrum from a time domain signal. The spectrum shows the magnitude of signal in frequency domain. If the bearing or gear has defects, the impact signal would produce abnormal components in the spectrum. According to the shape of these abnormal components and their frequencies, the experienced operator could identify that defect is BPFO, BPFI or gear teeth broken.

2.1.1 2.1.1 Fast Fourier transforms (FFT)

FFT is a popular and widely adopted tool for vibration based machine fault diagnosis. It constructs a frequency spectrum from the captured temporal signals. A typical FFT spectrum has been shown at the bottom window of Figure 1.

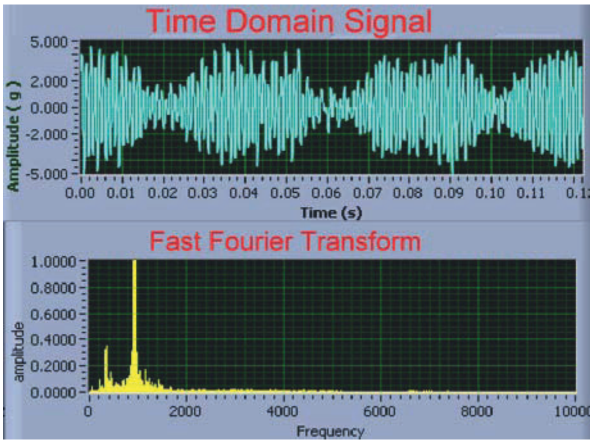


Fig. 1 the FFT spectrum shows the magnitudes of signals in the frequency domain.

If a component of the machine is defective, the magnitude of the vibration generated from this component will increase at its related excitation frequency. Comparison can be made to confirm the existence of fault by using the spectrum generated at the normal operating condition.

According to R.B. Randall, the function of Discrete Fourier Transform can be express as

$$X(k) = \sum_{n=1}^{N} x(n) * \exp(-j * 2 * pi * (k-1) * (n-1)/N),$$

for , 1 <= k <= N. Eq [1]

where N is the data length with vector x.

In Figure 2, the top diagram shows a defective bearing with a scratch occurring on the surface of the bearing outer-race. Every time when a ball or roller passes through the defective surface, an impact will occur. When the bearing is rotating at high speed, a series of impacts will be generated. Such impacts will create impulsive signals at high frequency region of the FFT spectrum as shown in the middle diagram of Figure 2. For comparison purpose, the frequency spectrum of a normal bearing is shown at the bottom diagram of Figure 2. Note that at high frequency region of the normal bearing's spectrum, there is no significant increase of amplitudes as compared to the one for defective bearing. During the test, since there were no other vibration generated from other nearby components, the results on the FFT spectra become obvious. Such results will allow the operator easily recognize whether a bearing is defective or not by only observing the comparison of the FFT spectra.

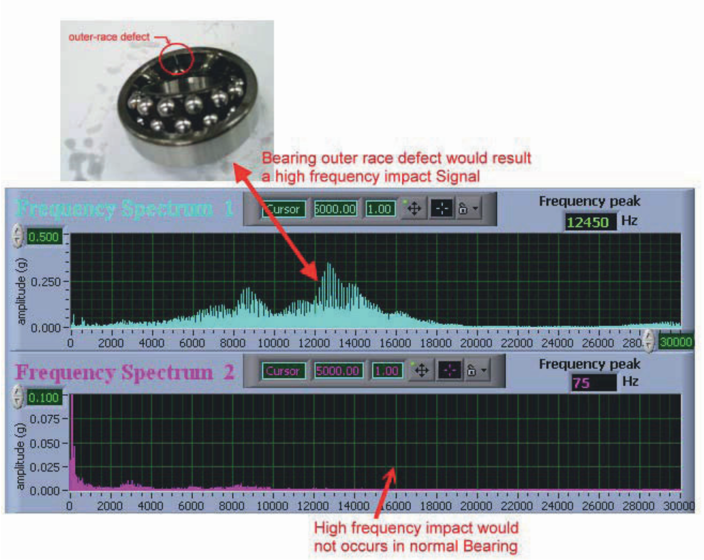


Fig 2. Difference in FFT spectra between a normal bearing and an outer-race defective bearing.

2.1.2 2.1.2 Vibration Trend Analysis

A vibration trend is a 3D plot that consists of many FFT spectra obtained at different time or days. As shown in Figure 3, the y-axis of the trend plot is the vibration amplitude, the x-axis is the Frequency, and the z-axis is the number of record obtained at different time. Since the trend plot shows the temporal information of vibration, it is helpful to observe the change of vibration in a particular range of frequency. If a component in a machine is deteriorating, its vibration level will be continuously increasing. Hence, by observing the increase of vibration level in the trend plot, one can determine the seriousness of a damaged component.

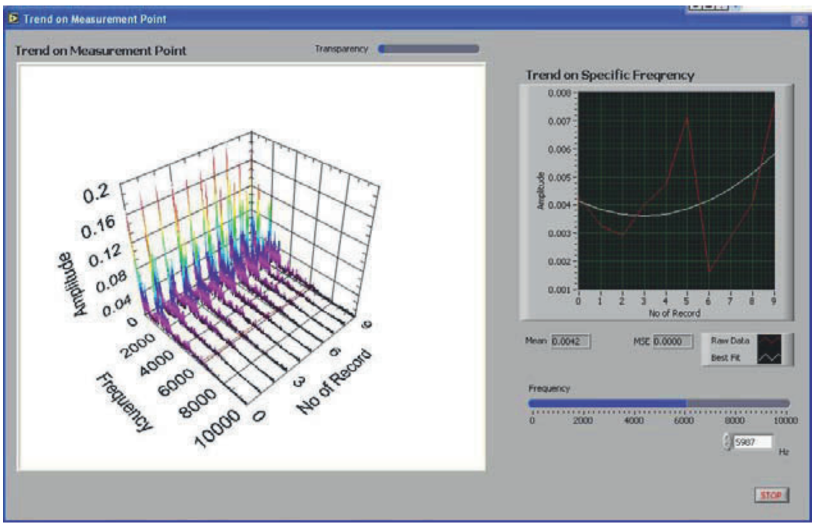


Fig. 3. The vibration trend analysis embedded in SAMS.

According to Larsen and Son, "the resulting trends curves can be extrapolated forward in time to indicate when the condition will reach a danger limit". If the operator wants to inspect the trend of a particular frequency range, he may specific that frequency range by clicking the frequency slide bar or input the desired frequency range in the given text box as shown on the right hand side of Figure 3. The vibration trend for the specific frequency range will be displayed on the right hand side graph. SAMS will provide a polynomial best-fit curve to fit the temporal values. Based on this curve, the operator can determine the future trend of the plot that related to the future deterioration rate of the damaged component. He then can prepare proper remedy in advanced.

2.1.3 Higher order statistical Analysis – Kurtosis (machine health indicator)

One mathematical representation of the deviation of an amplitude distribution from Gaussian is the so-called "fourth moment", or kurtosis. Kurtosis is a statistical indicator used to characterize the pulse character of a signal. It is a dimensionless parameter characteristic the flattening of the signal probability density. It corresponds to the 4th-order moment

The Gaussian distribution has a kurtosis of 3, and higher values of kurtosis indicate increased crest factor of the vibration signal. Kurtosis is a valid measure of the degradation of a machine, but it does not give any indication of the diagnosis of the problem. It has been reported that kurtosis is especially well suited to monitoring of reciprocating machines for fault detection.

For the data Y_1, Y_2... Y_N, the formula for kurtosis is:

$$Kurtosis = \frac{\sum_{i=1}^{n}(Yi - \bar{Y})^4}{(N-1)S^4} - 3 \qquad \text{Eq [2]}$$

$\bar{Y}$ is the mean,

S is the standard deviation,

N is the number of data points.

Because of the Kurtosis value for a normal distribution is three. Sometime we would minus three to reduce the Kurtosis value for normal distribution to become zero. For this reason, we change the representations of Kurtosis as below.

One possible advantage of using kurtosis as a fault detection parameter is that it does not need to be trended over time to be effective. A kurtosis of 3 generally is taken as indicating a healthy machine, with higher values indicating progressive states of fault progression. Besides, Kurtosis can be assimilated with a shape factor, the value of which does not depend on the signal amplitude:

Kurtosis is particularly suited to the monitoring of bearings of low speed rotating shafts, where frequency-based techniques are limited. Kurtosis is also widely used to detect non periodic shocks. In practice, Kurtosis is often calculated after filtering. It can reach values greater than 100, especially when the selected frequency band coincides with a structure resonance. Also, it is often associated with envelope analysis to determine the area to demodulate.

2.2 Enhanced fault diagnosis methodology

2.2.1 2.2.1 Continuous Wavelet Transform (CWT)

According to Strang G., and Nguyen T, the wavelet transform can be expressed as:

$$C_{(a,b)} = \int_{-\infty}^{\infty} f(x) Wa, b(x) \partial x$$

$$\text{where } W_{a,b}(x) = \frac{1}{\sqrt{a}} W\left(\frac{x-b}{a}\right) \qquad \text{Eq [3]}$$

From Tse and Yang, the machine fault signals are always buried with other irrelevant signals due to the problems of overlapping and energy leakage. FFT may not easily detect the defect due to the abandonment of the time information. In SAMS, both wavelet transform related techniques - CWT and EWA, have been developed to enhance the ability in machine fault diagnosis. Particularly, the EWA can minimize the problems of overlapping and energy leakage that always occur in the results of conventional wavelet transforms. For wavelet transforms, SAMS provides Mexican Hat, Mayer, Morlet, Coif, and Daubechies wavelets. Their parameters are manually adjustable to suit the analyses for different type of signal with different characteristics. The diagnosing results by applying the CWT to analyse the vibration signals generated by a normal bearing and a defective bearing are shown in Figure 4. The bright colour spots and strips as shown in the right hand side diagram of Figure 4 reveal the high-energy impulses caused by the defective bearing. They form an advanced warning for serious faults may occur if the defective bearing is allowed to be deteriorated. Note that the CWT provides a time-frequency analysis simultaneously that the conventional FFT spectrum fails to provide.

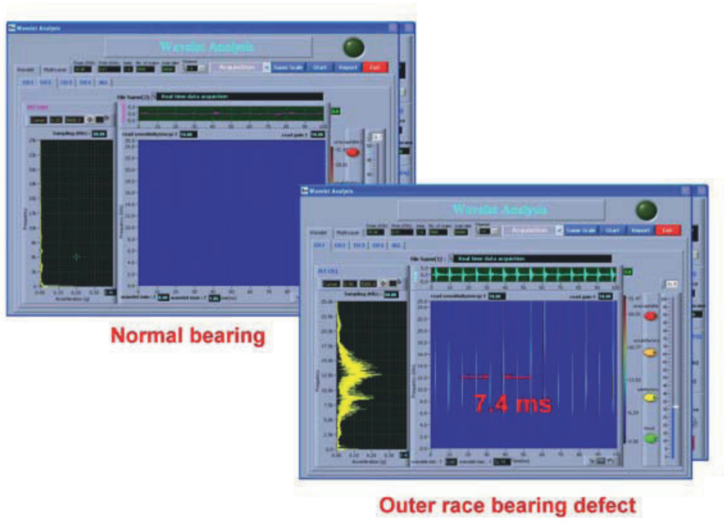


Fig. 4. a captured bearing vibration signal, with outer race defect, under Wavelet analysis.

Using the common equations for obtaining the bearing characteristic Frequencies, the outer-race defect frequency, f_{od}, generated by a bearing with known specifications can be calculated as:

$$f_{od} = 0.5 * (1 - \gamma) Z f_{s,} = 5.8\, f_s = 135 \text{ Hz (7.4ms)} \qquad \text{Eq [4]}$$

where Z is the number of rolling elements, f_s is the fundamental rotational frequency, and γ is a parameter related to the dimensions of the bearing and its contact angle.

Given that the inspected bearing is SKF 1206 EKTN9, the calculated outer-race defect frequency should be 135 Hz or about 7.4 ms. Note that in the right hand side diagram of Figure 4, the impacts occur at a time interval of 7.4ms, which is similar to the calculated value of the outer-race defect frequency. Hence, the cause of impacts is an outer-race bearing defect. To conclude, with the help of the time-frequency analysis offered by CWT, one can easily observe the faulty impacts and identify the cause of defect.

Another advantage to use CWT in machine fault diagnosis is the CWT can provide signal decomposition, which is especially suitable for identifying the excitation frequency generated by each component. Note that in the right hand side diagram of Figure 4, the excitation frequency of the bearing housing is peaked at around 12.5 kHz. Since, different structure should have different excitation frequency, hence, the vibration generated by each location of the machine could be revealed at its related excitation frequency. By observing the change of each excitation frequency, the health of the component at that particular location can be determined. Therefore, the signal decomposition plot is useful in identifying the source of severe vibration.

2.2.2 Active Noise Cancellation (ANC)

Although the CWT and EWA can provide time-frequency analysis to enhance the accuracy in fault identification, they work better if the inspected machine is large and each component is well apart from each other. If the inspected machine is small with vibration generated by each component interfering each other, then a more advanced technique should be employed to minimize the inference. If the interference has not been minimized, the results generated by any technique will be very difficult to be interpreted. Inexperienced operator or maintenance staff may misinterpret the results and false alarms may occur. To minimize these undesirable effects, the SAMS has embedded with the virtual instrument called active noise cancellation (ANC). ANC is capable of minimizing interference as well as optimizing the desired signal. SAMS has provided two types of ANC for different need. The first type of ANC is for known interference, that is, the operator knows which source of signal should be minimized. Its block diagram is shown in Figure 5. The primary signal should contain the desired signal and the interference, whilst the reference signal should contain only the undesirable interference. The output signal will mainly compose of the desired signal.

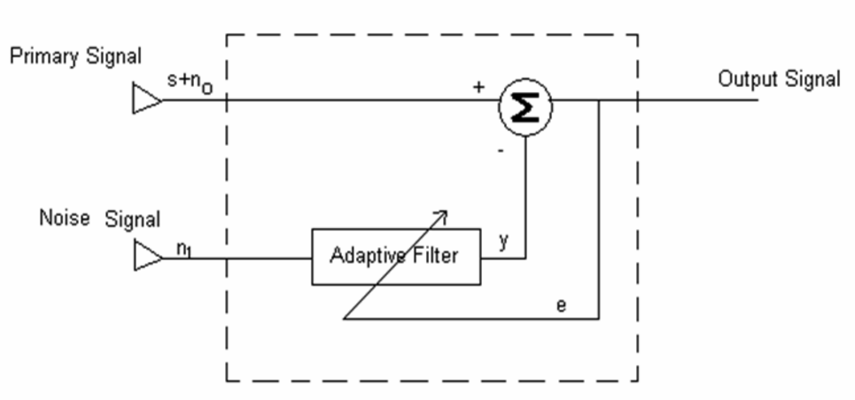


Fig. 5. The first type of ANC for canceling interference.

Sometimes, the interference cannot be identified, but the desired signal can be simulated, likes the classical Model Reference Control. Therefore, the first type of ANC must be modified to suit for inputting the desired reference instead of the interference. In such case, the second type of ANC has been developed. Its block diagram is shown in Figure 6. This time, the primary signal is also the aggregation of the desired signal and the interference, whilst the reference signal should contain only the simulated desired signal. After the applying of ANC, the output signal will mainly compose of the real desired signal.

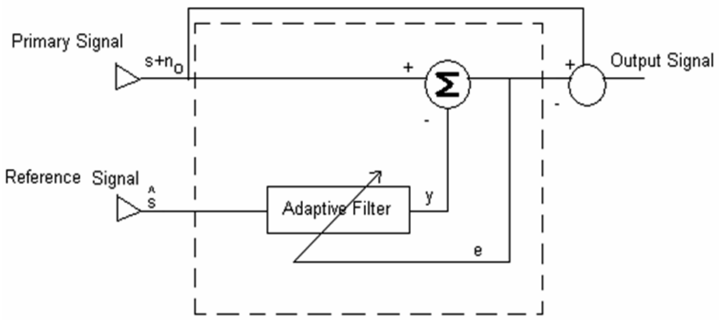


Fig. 6. The second type of ANC with feed-forward loop.

With the help of ANC, the faulty signal becomes easier to be revealed and false alarms can be avoided. To verify the effectiveness of the first type of ANC, a series of tests using microphones to capture the vibro-acoustic signal generated from a roller of a print press were conducted in a printing factory. As shown in Figure 7, two microphones were used. One microphone was pointing at the inspected roller. Another microphone was pointing at the surrounding environment. The sound collected by the microphone pointed to the roller is shown at the top diagram of Figure 7. The sound collected by the microphone pointed to the surrounding environment roller is shown at the middle diagram of Figure 7. Note that by comparing the top and the middle diagrams, the first sound has been modulated by the surrounding sound. By using the first type of ANC, the surrounding sound has been minimized. Hence, the vibro-acoustic signal from the roller can be recovered as shown in the bottom diagram of Figure 7. Therefore, the fault related signals are easier to be revealed in after the application of ANC. Here, at least two microphones have to be used to make ANC becomes effective.

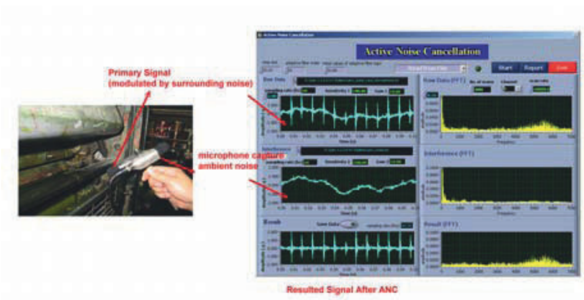


Fig. 7: The recovered faulty signal after applying ANC.

2.2.3 Orbit Analysis

Orbit Analysis is useful to check misalignment and unbalancing occurred in all rotating shafts. The orbit plot is constructed by combining the displacement signals collected by a horizontal sensor (in X direction) and a vertical sensor (in Y direction). In SAMS, we calculate the displacements by integrating twice the acceleration signals captured by the two sensors or accelerometers. The reason to obtain displacement signals from acceleration signals is to save the extra cost of buying displacement sensors as accelerometers are already existing and needed for other fault diagnostic techniques. Beside integration, the processes of de-trending and filtering out other signals except the signal related to the fundamental rotational frequency must be employed to obtain a clear track of orbit.

Figure 8 shows two orbit plots, one from a balanced shaft and another one from an unbalanced shaft. Generally speaking, a balance shaft will produce a much smaller and round orbit. Whilst, an unbalance shaft will produce a relatively bigger orbit with irregular shape biased to a certain angle. Note that the SAMS' orbit analysis is very sensitive to problem shaft. Even a slight misalignment or unbalancing will generate a substantial large orbit as compare to a normal orbit.

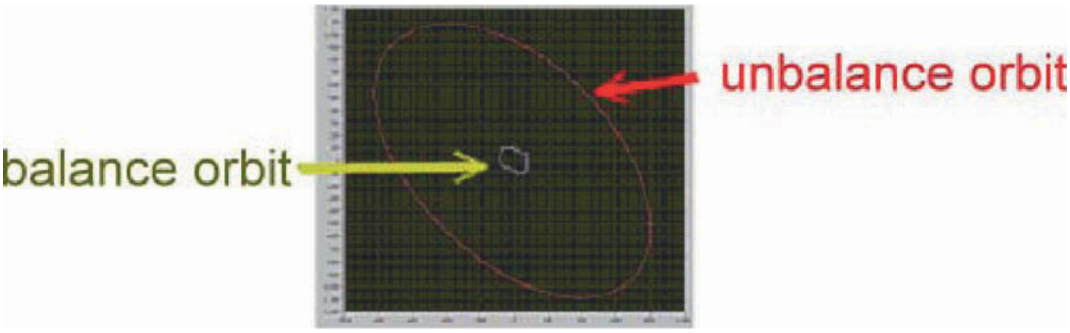


Fig. 8. Unbalance orbit is obvious in orbit analysis.

3 CONFIGURATION OF SMART ASSET MAINTENANCE SYSTEM (SAMS)

Table 1	
The device used for capturing the sound and vibration signal.	
	1. Accelerometer - a wide range of frequency and high sensitivity accelerometer is required to detect the vibration signal.
	2. Microphone - a pre-polarized free field and wide frequency range microphone is used to detect the sound signal.
	3. Signal conditioner - provide amplification and filtering of analogy signal. It will provide power to ICP type of sensors.
	4. Data Acquisition - capture and transform the analogy signal to digital.
	5. Portable Computer – a notebook PC or a PDA for software installation and data recording.

3.1 Instruments used for conditional signal acquisition

Refer to the introduction; vibration is normally a destructive by-product of the force transmission through a machine. Also, vibration carries much information relating to the machine health. Therefore, sensing of the vibration is important for maintenance. To acquire the vibration or vibro-acoustic signals generated from an inspected machine, a number of devices must be employed, a portable vibration analyser set and level recorder is excellent. Table 1 shows all the devices that are currently cooperated with SAMS.

For vibration type of machine fault diagnosis, we use the ICP accelerometers and microphones to acquire vibration and vibro-acoustic sensory signals respectively. For other purposes, such as acquiring pressure, temperature, current, voltage etc., SAMS has built-in functions to collect these data. Here, we only concentrate on the description of using SAMS for vibration type of machine fault diagnosis. The signals collected by the sensors will be conditioned by a multi-channel signal conditioner prior to input to a data acquisition (DAQ) card for capturing and converting conditioned signals to digital format ready to be accepted by a notebook computer or a PDA. Figure 9 shows the overall instruments required by SAMS from capturing signals by a sensor to a computer embedded with a DAQ card. In short, the computer becomes a physical analyzer for vibration monitoring and fault diagnosis.

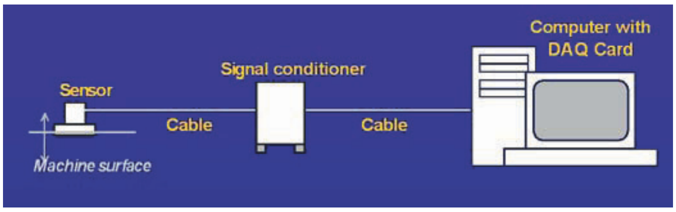


Fig. 9. The overall instruments work with SAMS.

Usually accelerometers are used for collecting vibration generated from a machine as long as the accelerometers can be mounted on the desire locations on the machine's surface. However, a lot of components of the machine may be moving or difficult to be accessed so that mounting the accelerometers on the components' surfaces is not feasible. In such case, then microphones may be used to replace accelerometers as they are non-contact to the surfaces. Due to high background noise in a factory or plant, using one microphone for collecting vibro-acoustic signals will definitely have a very low signal-to-noise ratio. Hence, advanced methods, such as ANC and BSS must be employed to separate noise from real vibration signal. Usually, these advanced methods may require at least two microphones. By using the method of microphones to collect signals, it has significantly increased the flexibility in data collection that has encountered access difficulty.

3.2 Selected description on SAMS 's function

SAMS is acting like a portable virtual analyser for machine monitoring and fault diagnosis. Its platform provides a graphical user interface (GUI) that allows the user to choose different virtual instruments for different diagnostic techniques. Besides, SAMS contains a comprehensive set of tools for acquiring, analysing, displaying, publishing and reporting, as well as a database for the storage and retrieval of the acquired machine data. Figure 10 shows the GUI menu of SAMS.

Fig. 10. The graphical user interface of SAMS.

The functions of SAMS include five main areas, 1) the DAQ and signal processing; 2) the data base manager; 3) the fault analysis and trend prediction; 4) the automatic report generation, and 5) remote monitoring via the Internet.

In the panel of DAQ and signal processing (see Figure 10), the input signals are sampled in accordance with the prescribed sampling rate, number of samples, sensor sensitivity and its gain. Once the prescribed number of data has been acquired, the sampled temporal waveforms will be displayed in one of the four channels as shown in Figure 10. The data collection process can be continuous and all four waveform plots will be updated momentarily. Statistical values of the captured signals, such as RMS, peak, mean, standard deviation, Kurtosis, skewness, and crest factor etc. will be determined and displayed here for reference.

Four LED indicators are included on the right hand side of each temporal waveform plot. Each LED indicator is intended to be lighted up according to its associated energy level contained in the vibration signal captured by each of the four channels. Different colours, like the green, yellow, orange and red colour have been assigned to each LED to represent good, satisfactory, unsatisfactory, and unacceptable vibration levels respectively. Simply by observing which colour of the LED has been lightened, the machine operator can instantaneously determine the health condition of the monitoring machine.

A Digital Filtering menu is also included in this panel. The digital filters are available for the removal of undesirable frequency components in the acquired signals. A set of selectable high-pass and low-pass filters are available for such purpose. Scaling of the amplitude axes can be made instantly by simply typing the required scale into the Axis Scale control box. The machine operator has all freedom to choose his desirable scale for each display axis. Moreover, integration buttons for all channels are installed so that the operator can collect the vibration signals in terms of acceleration, velocity, or displacement as desired.

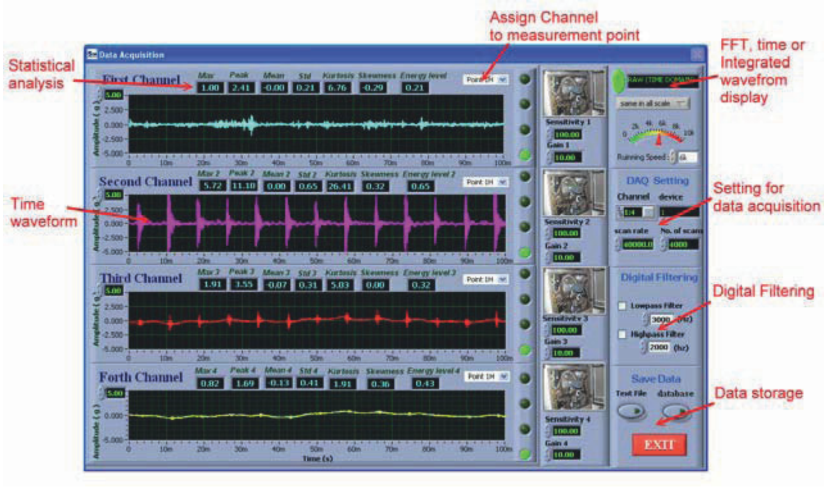


Fig. 11 the Data acquisition and signal processing panel.

To enable the operator to analyse vibration data, a database with data management facility has been developed in SAMS. As shown in Figure 12, the hierarchy of the database consists of 5 entities, namely, the company identification, the machine identification, the inspecting component, the measurement location and orientation, and the data. To store the sensory data into the database, the user must first define the properties of the company, machine, component and measurement in the database. For a company, its properties could be its address and telephone number, whilst, for a machine, its properties could be its power and running speed. In addition, a visual data explorer (data manager) is included in SAMS. The data manager will keep tracking all the data records that have been collected according to their properties automatically.

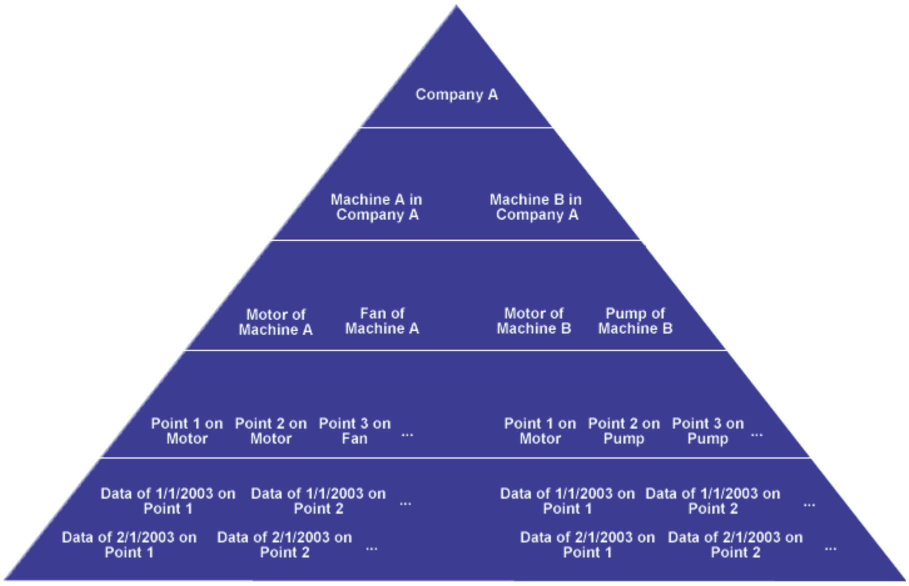


Fig. 12. The database construction in SAMS

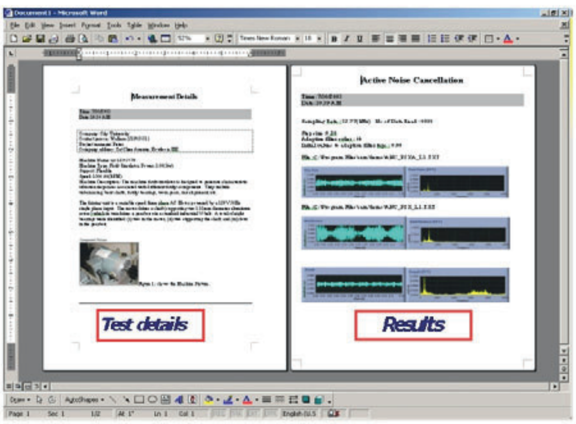


Fig. 13. Automatic generation of analysis report in Word format.

In commercially available machine monitoring systems, after the data acquisition and the fault diagnosis processes have been completed, operators or experts have to spend a lot of time on reporting and documentation. To ease such works, SAMS has a built-in function for generating reports and documents simply by pressing a button. On each page of panel or virtual instrument, there is a button called “Report”. Once the button is pressed, the displayed window with its analysing results, related sampling data, and details of the inspected machine etc., will be automatically inserted into a formatted Word document as shown in Figure 13. This can not only save the documentation time, but also ensure all the technical and physical references are in the generated document or report. Since the document is in Microsoft Word format, it is very convenience for the operator or expert to further edit the file according to his preference.

4 SELECTED INDUSTRIAL APPLICATIONS

SAMS has been tested on many industrial machines. A list of tested machines includes traction motors on trains, chillers in buildings, water pumps, stacker cranes for logistics and warehouse, printing press and associated machines, beverage production lines, coal milling machines, catering production line etc. Over 20 companies that belonged to the sectors of public utilities, mass transportation, logistics, and large manufacturing has been using SAMS on their equipment monitoring and fault diagnosis. Figure 15 shows some of the industrial machines that have been tested by SAMS. Due to the limited length of this paper, only a few selected applications are highlighted here.

Fig. 14. Some industrial machines tested by SAMS.

5 CONCLUSIONS

In this essay, an innovative solution for equipment health monitoring called, the smart asset maintenance system (SAMS) has been introduced. With the help of the advanced virtual instruments embedded in SAMS and its powerful methodology, the machine operators can easily detect the occurrence of faults directly from the machine-operating signal. Besides, SAMS contains many comprehensive functions for acquiring and analysing machine running signals, displaying the fault diagnostic results, publishing the results and acquired data on Internet, predicting the deterioration rate of a defective component, automatically generating reports, as well as forming a user-friendly database for the storage and retrieval of the acquired machine data. SAMS has been tested with many industrial machines owned by more than 30 companies, including companies in public utilities, mass transportation, logistics, and large manufacturing. Currently, SAMS is equipped with wireless communication ability that enables a PDA to collect machine data wirelessly. With so many advanced functions built-in SAMS, the current cost of SAMS is less than US$2,000 per system. Hence, SAMS is low-cost yet equipped with the foremost functions for machine health monitoring and its degradation prediction. For continuous developments, the future SAMS will allow not only wirelessly collecting signal, but also wirelessly controlling the monitoring equipment. Moreover, the health information of the inspected machine can be transmitted to remote user via SMS and GPRS that are currently available on smart mobile phones.

ACKNOWLEGMENTS

The work described in this paper was partly supported by a grant from the Innovation and Technology Commission of the Hong Kong Special Administrative Region, China (Project No. ITSP/169/01) and partly supported by a grant from the City University of Hong Kong, China (project no. 7001808).

6 REFERENCE LIST

[1] Ashok Ambardar, "Analogy and Digital Signal Processing", Michigan Technological University

[2] Donald E. Bently., Charles T. Hatch and Bob Grissom., "Fundamentals of Rotating Machinery Diagnostics"., Bently Pressurized Bearing Company, 2002

[3] Daniel J. Inman., "Engineering vibration"., 2nd edition., Prentice Hall., 2001.Ramirez,

[4] K. Larsen & Son, "Machine health monitoring using vibration analysis", Bruel and Kjaer

[5] Larsen K., and Son A., *Machines Condition Monitoring*, 1989, Bruel & Kjaer, November, p.13. Randall R.B., *Frequency Analysis*, 1987, Bruel and Kjaer Technical Review, 3rd edition, Sept, p. 28.

[6] R.B Randall, B. Tech., B.A., Frequency Analysis., 3rd edition., 1987

[7] Strang G., and Nguyen T., *Wavelets and Filter Banks*, 1996, Wellesley Cambridge, p. 432.

[8] Tse P., Yang W., and Tam H., "Machine Fault Diagnosis Through an Effective Exact Wavelet Analysis", *Journal of Sound and Vibration,* Vol. 277(4-5), Nov. 5, 2004, pp.1005-1024.

[9] Tse P., and Lai W., "Minimizing interference in bearing condition monitoring and diagnosis by using adaptive noise cancellation", *Proceedings of the 16thInternational Congress and Exhibition on Condition Monitoring and Diagnostic Engineering Management,* Vaxjo, Sweden, August 2003, pp. 739-748.

[10] Wang J., Tse P., He L., Yeung R., "Remote Sensing, Diagnosis and Collaborative Maintenance with Web-enabled Virtual Instruments and Mini-Servers", *International Journal of Advanced Manufacturing Technology*, Vol. 24(9-10), Nov. 2004, pp. 764-772.

Glossary

ICP – Integrated circuit piezoelectric

SECURITY CONSIDERATIONS FOR MODERN WEB-BASED MAINTENANCE OR REMOTE SENSING SYSTEM

Alan N. Cheuk and Peter W. Tse

Smart Engineering Asset Management Laboratory (SEAM) Department of Manufacturing Engineering and Engineering Management City University of Hong Kong, Tat Chee Avenue, Hong Kong, P. R., China E-mail: meptse@cityu.edu.hk Tel: (852) 27888431 Fax: (852) 27888423

Abstract: In an ever-increasing competitive environment, companies must maintain the quality of their production and services. Hence, machinery or equipment that is critical to production and providing services must be maintained with minimum downtime or a near zero-break down rate. The concept of Web-based maintenance helps to detect faulty equipment at an early stage without the constraints of distance. It provides a convenient platform for maintenance personnel to evaluate the performance of machines even when they are distant. As the Internet is an open environment, its traffic or on-line communication can be easily intercepted. Users who do not adequately protect their Internet communication are exposed to risk. This paper intordouces a Web-based maintenance system with security considerations. With the assistance of several security measures, the data security of our Web-based maintenance system has been enhanced. The data can be transferred safely and the accuracy of the results can be verified. Our Web-based maintenance system works more securely in comparison to conventional Web-based maintenance systems that have been reported thus far.

Key Words: Web-based maintenance; Remote sensing; Data encryption; Internet security.

1 INTRODUCTION

Manufacturing industries are currently faced with the challenge of reducing manufacturing costs but maintaining high productivity and customer satisfaction. An effective maintenance strategy is thus important to survival in a competitive environment. A Web-based maintenance system allows operators to remotely monitor equipment and perform fault analysis. Through the Internet, sensory information and the operating conditions of various types of equipment can be transferred from a factory, for example, to a remotely located office. The managers or engineers can evaluate the working condition of equipment using built-in and Web-enabled virtual instruments that are provided by the server via the Internet. Hence, the cost of maintenance and labour can be reduced, and unscheduled downtime can be minimised. Internet use creates security concerns. While the Internet provides a convenient medium for performing remote monitoring and analysis, the security risk cannot be ignored. As the Internet is as an open medium, hackers can enter Internet channels and intercept information. They can then divert or alter the intercepted information for their own advantage. To deal with the problem of data security, cryptography is a solution. In our Web-based maintenance system, several security mechanisms have been put in place to protect the privacy of data. Confidentiality, authentication, and message integrity are usually the key components of secure communication [1]. Each of these performs a unique role in securing information systems. Data confidentiality is the protection of data from unauthorised disclosure. It protects data from passive threats. Authentication is the protection of data by proving identity. On the Internet, everyone is anonymous. It should be possible for the receiver of a message to ascertain its origin. Integrity is the protection of data from tampering or forgery. The receiver should be able to verify that it has not been modified in transit. By ensuring data confidentiality, integrity, and authentication, machine data can be transferred more safely, which may lead to accurate machine fault identification.

2 SECURITY TECHNOLOGIES

Cryptography is the art and science of keeping messages secure. It helps to prevent intruders from being able to use the information that they intercept. Maintaining data confidentiality, integrity, and authentication are the main issues of protection. The design and implementation of the security mechanisms that are used in our system can be described in three ways: the encryption, message digest, and digital signature. Confidentiality is achieved by encryption, message integrity is accomplished by message digest and authentication is attained by digital signature.

Encryption is the process of changing information (plaintext) into a format (ciphertext) that is unreadable or unusable [2] and [3]. It can be classified into two categories: symmetric and asymmetric. A symmetric algorithm applies the same key for encryption and decryption. Once the key is agreed upon by two communicating parties, it is known by only the sender and the receiver. No one else can access it. An asymmetric algorithm applies different keys (a public key and a private key) for encryption and decryption [4]. A public key is available to everyone, which means that anyone can perform encryption and signature validation. A private key is undisclosed and is used to decrypt messages that have been encrypted by a corresponding public key. A sender encrypts a message with the recipient's public key and sends the encrypted message to the recipient. Only the recipient who knows the corresponding private key can decrypt the message [5]. A symmetric algorithm is an efficient encryption method that can be performed quickly. However, there is a problem with key distribution [c]. An asymmetric algorithm can solve the problem of key distribution, but it has the drawback of slow performance. An alternative solution is to use both symmetric and asymmetric algorithms [6] and [7]. Thus, both symmetric and asymmetric algorithms are applied to our Web-based maintenance system to achieve both advantages. An asymmetric algorithm is used to exchange the undisclosed secret key that is used in the symmetric algorithm. Subsequently, a symmetric algorithm is used to exchange bulky machine signals. This can combine the fast performance of a symmetric algorithm and the key distribution of an asymmetric algorithm.

To achieve reasonable security in data transmission, ensuring confidentiality, integrity, and authenticity are needed [b]. However, the use of encryption can only accomplish confidentiality. Hackers can still alter the encrypted message by adding or deleting some of the message content. Moreover, a recipient who obtains the key can only decrypt the message – it is difficult to determine the origin of the message, and the received message may not originate from where the corresponding party claims that it originates. Hence, other security mechanisms are needed to achieve data integrity and authenticity. The combination of encryption, message digest and digital signature is the solution. Message digest is the primary mechanism in achieving message integrity [8]. It is obtained by a hash function, which is a mathematical formula that converts an arbitrary-length input (message) into a fixed-length output (hash value) [9]. The hash value is transferred with the original message to the recipient. After decryption, a recipient carries out the same hash function in relation to the message. We can verify that the received message is not altered if the hash value is the same as the one received. Otherwise, the message has been altered or compromised during transmission. An essential property of hash function is that it is a one-way function. If someone obtains the hash value of a hash function he or she cannot reverse the hash function to retrieve the input. A further property of the hash function is it is collision free. It is assumed that we cannot create the same hash value from two different inputs, while little change of the input will result in a very different hash value. This characteristic helps to increase the difficulty of accessing the original message. While hashing can provide data integrity, it is ineffective if it does not contain origin authentication [10]. When receiving a hash value and its origin message, it is difficult to verify that the message originates from the identified sender. Hence, hashing is needed to include a proof of the sender's identity. A digital signature is needed to provide message authenticity and prevent tampering [d]. A digital signature is a combination of hash function and an asymmetric encryption. A message is first subjected to a hash function. The hash value is then encrypted with the sender's private key to form a digital signature. Once it is received, the receiver can use the sender's public key to decrypt the digital signature. As the public key is the only key with which to decrypt the message that has been encrypted by its corresponding private key, and as the sender's public key can decrypt the digital signature, we can verify whether the message originates from the identified sender.

3 APPLICATION OF SECURITY MECHANISMS IN WEB-BASED MAINTENANCE

The architecture of our Web-based maintenance system is shown in Figure 1. This system can be divided into three main components: local items, server, and remote user. A local computer is used to carry out data acquisition and data publishing. A server is where the analysis and storage of diagnostic programs are performed. To perform a diagnostic test, the server first collects the data from the local computer. At the same time, the collected data and the required diagnostic program are sent to the requested user for analysis. Yet, the Internet is an insecure environment. The security risk cannot be ignored. Figure 2 shows a possible path that a hacker may intercept or through which he or she may alter a machine's signal that is transmitted via the Internet. Consider the case in which a maintenance person wishes to perform an analysis to check the performance of a remotely located machine. The maintenance person must first collect the raw signal of the inspected machine. Once the data has left its origin and has been transferred via the Internet, there is the possibility of a hacker diverting or altering the raw data. The altered raw data thus affects the accuracy of the analysis, which is performed by the maintenance person at a separate location.

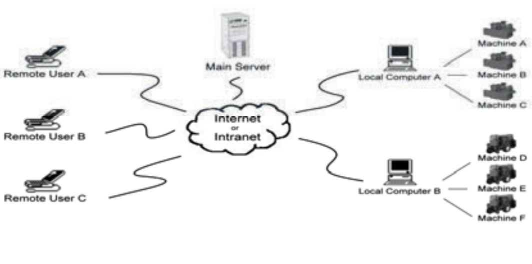


Figure 1: Architecture of Web-based maintenance system

The data protection processes that are used in our Web-based maintenance system are formed by combining symmetric encryption, asymmetric encryption, hashing, and digital signature. As asymmetric encryption does not contain the problem of key distribution, it is used for key exchange. The secret key that is used in symmetric encryption is first exchanged with the aid of asymmetric encryption. Subsequently, bulky machine data can be protected by symmetric encryption that can be performed quickly. This alternative approach can utilise both the advantages of symmetric (fast performance) and asymmetric (key exchange) encryption. The diagram on the left of Figure 3 shows the overall key exchange processes that are used in our system. The secret key is first transferred to a hash function. The hash value is then subject to an asymmetric encryption with the sender's private key to create a digital signature. This digital signature and the original secret key are then combined and transferred to an asymmetric encryption by a receiver's public key to form the final encrypted key block. Once it is obtained, the receiver can decrypt the encrypted key block with a private key to retrieve the secret key and the digital signature. As the sender's public key is openly accessible, the receiver can easily obtain the public key of the sender to decrypt the digital signature. To verify the secret key, it is transferred to the same hash function that is performed by the sender and the result is compared with the one that is obtained in the digital signature. If the two values are exactly the same, then we can verify that the secret key is valid and that it is sent from the local computer it claims to have been sent from.

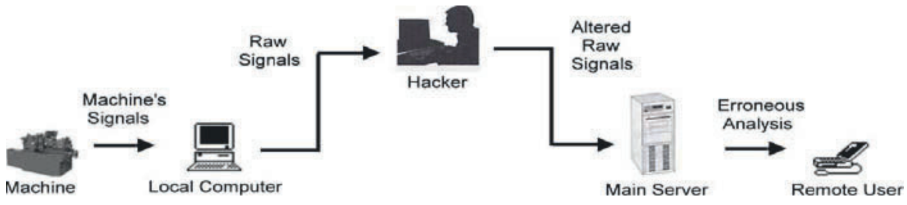


Figure 2: Data security problem in Web-based maintenance system

After the secret key exchange, symmetric encryption is used for further data protection. The right diagram in Figure 3 shows the overall data exchange processes that are used in our system. The operation of data exchange is similar to key exchange. The only difference is that the algorithm that is used for data encryption is changed to a symmetric encryption. Data is first transferred to a hash function to obtain a hash value. This hash value is then encrypted (asymmetric algorithm) with the sender's private key to create a digital signature. The original data and the corresponding digital signature concatenate and carry out a symmetric encryption to create the final encrypted data block. The encrypted data is unreadable by other parties. To retrieve the original data and the corresponding digital signature, the receiver uses the secret key that has already been obtained to decrypt the encrypted data block. The decrypted data is then transferred to the same hash function to produce a new hash value for comparison. If this hash value is the same as the one that is obtained in the digital signature, then the data can be verified as valid and can be used for diagnostic analysis.

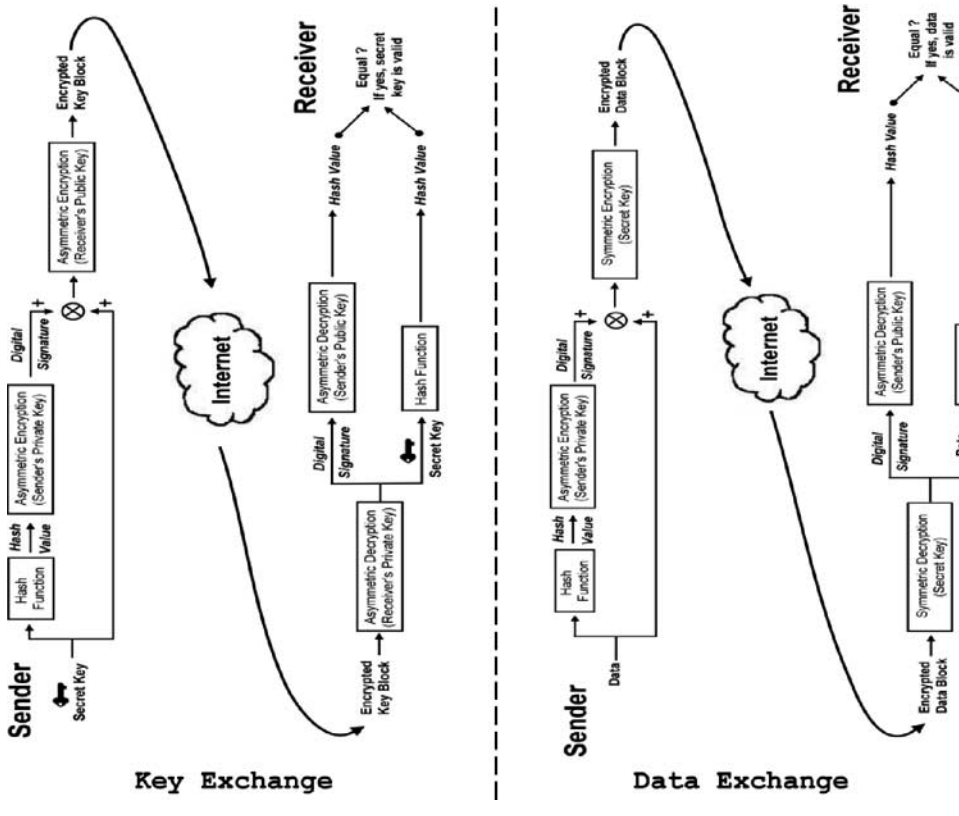


Figure 3: Security processes for key exchange and data exchange

4 SECURITY ALGORITHMS USED IN WEB-BASED MAINTENANCE

The algorithms that are used in our system for symmetric encryption are the Data Encryption Standards (DES) and the Advanced Encryption Standard (AES) [11] and [12]. They are Federal Information Processing Standards (FIPS) that are managed by the National Institute of Standards and Technology (NIST). The NIST is a federal technology agency in the United States that develops and promotes measurement, standards, and technology. The Federal Information Processing Standards Publication (FIPS PUB) is developed by NIST as guidelines for communicating and securing data in information systems. DES is designed for enciphering and deciphering blocks of 64-bits. The key in DES is also 64-bits, but only 56-bits are used (remaining 8-bits are used for parity checks). AES is designed for enciphering and deciphering blocks of 128 bits. The key in AES can have three sizes: 128-bits, 192-bits, and 256-bits. For the point of security, AES is stronger than DES as it is the replacement of DES approved by the US NIST. While DES is less secure than AES, it still has its value in that the processing time of DES is shorter than AES. Thus, our system has implemented both DES and AES as an option for the system user. Users need to determine which they are more concerned about, processing time or protection level, and to select their favourite algorithms.

The algorithm RSA is used for asymmetric encryption. RSA is derived from the first letters of the surnames of its creators (Rivest, Shamir and Adleman). In our Web-based maintenance system, RSA is used to protect the secret key that is used in DES or AES and to produce the digital signature. The basic operation of RSA depends on the factoring problem. Factoring problem is the problem of finding which factors are multiplied to produce a particular number. It is straightforward to multiply several numbers, but it is not easy to figure out what numbers were combined to produced a particular result [13]. RSA contains three main parts: public key e, private key d, and the shared value N (which is often called modulus). The public key e_r and modulus N_r of the receiver are open to the sender. The private key d_r remains undisclosed so that only the receiver knows its value. To encrypt the secret key, the sender makes use of the receiver's public key e_r and modulus N_r. To decrypt it, the receiver can make use of their own private key d_r and modulus N_r. To produce a digital signature, the sender uses their own private key d_s and modulus N_s to encrypt the hash value of the message. The sender's public key e_s and modulus N_s are the components that verify the authentication of the message.

The hash function that is used in our Web-based maintenance system is the Secure Hash Algorithm SHA-1 [14]. SHA-1 conforms to Federal Information Processing Standards (FIPS) for data security and is used in a large variety of security applications to ensure message integrity. Essentially, SHA-1 is similar in structure to another common hash function that is called MD5. The reason for implementing SHA-1, but not MD5, is that SHA-1 produces a larger message digest than MD5. The message digest of MD5 is 128-bits (approximately 3.4 x 10^{38} possible hash values), while the message digest of SHA-1 is 160-bits (approximately 1.4 x 10^{48} possible hash values). As SHA-1 produces a larger message digest, it is less likely to produce the same message digest with two different inputs. Due to the difference in message digest size, SHA-1 seems to be more suitable for our application. Although SHA-1 is 35% slower than MD5, it is a reasonable trade-off to obtain a better security performance. Hence, SHA-1 is selected as the hashing algorithm of our Web-based maintenance system.

In addition to the above algorithms, our Web-based maintenance system also provides another type of protection. As the diagnostic program is sent to remote users to perform diagnostic tests, the first thing we do is restrict the access right of the remote user (viewing, control, or deny). Each remote user is identified by a personal IP address and is given access right. For a remote user who only has viewing privilege, he or she can only view the analysed result in the program but cannot control any function in it. This can prevent an inexperienced user from inputting an invalid parameter into the system. Furthermore, the server can refuse the requests of particular remote users. This helps to prevent overuse of the system. Although unauthorised users can change their IP addresses to pose as authorised users, they must still know the value of a valid IP address. That is, unauthorised users cannot obtain access rights until a valid IP address is matched. The second protection is to restrict the usability of the diagnostic program. As all of the diagnostic programs are stored in the server, it is necessary to specify which programs are accessible to remote users. The server only publishes programs that can be accessed by authorised remote user. Even remote users who satisfy the above requirements (valid access right and valid program) must provide the correct password. To obtain the diagnostic program, remote users must input a URL to link up with the server. Before obtaining the required diagnostic tools, they must input a correct password to prove their identities. Otherwise, they are unable to obtain a diagnostic program.

5 CONCLUSIONS

This paper shows the architecture and the security processes of our Web-based maintenance. With Internet use and the advanced function of diagnostic programs, authorised managers or engineers can determine if machines are faulty without having to be on-site. Security is a prime concern for every Web-based system. To ensure the security of a machine's communication, several security mechanisms have been added to protect the privacy of data communication. The combined use of symmetric encryption and asymmetric encryption provides better encryption performance. The hash function and digital signature facilitate and ensure data integrity and authenticity. In addition, restricting the access right and usability of the program helps to prevent illegal use of our system. Coupled with password protection, Web-based maintenance can achieve reasonable security in data communication. By using our Web-based maintenance, sensory information and the operating conditions of various types of machines can be securely transferred from a factory, for example, to a remotely located office. Managers or engineers can then quickly evaluate the working condition of the equipment. As a result, companies can maximise their manufacturing productivity and customer satisfaction.

6 REFERENCES

[1] Kurose J.F., & Ross K.W. (2003) Computer Networking A top-Down Approach Featuring the Internet. Addison Wesley

[2] Champlain J.J. (2003) Auditing Information Systems. New Jersey, CA: John Wiley & Sons Inc.

[3] Ogletree T.W. (2004) Upgrading and Repairing Networks 4th Edition. Indiana, Que Publishing.

[4] Speed T., Ellis J. & Korper S. (2002) The Personal Internet Security Guidebook. San Diego, Academic Press.

[5] Klobučar T. & Jerman-Blažič B. (2001) Certificate Polices and Certification Practice Statements. Advanced Security Technologies in Networking. pp. 63-73. Amsterdam: IOS Press.

[6] Arazi B. (1993) Integrating a Key Distribution Procedure into the Digital Signature Standard. Electronics Letters, Volume 29, pp.966-967.

[7] Jerman-Blažič B., Schneider W. and Schwiderski-Grosche S. (2001) Introduction to IT-Security in Open Systems. Advanced Security Technologies in Networking, pp. 3-19, Amsterdam: IOS Press.

[8] Lail B.M. (2002) Broadband Network & Device Security. California, McGraw-Hill/Osborne.

[9] Clarke D., Devadas S., Dijk M.V., Gassend B. and Suh G.E. (2003) Incremental Multiset Hash Function and Their Application to Memory Integrity Checking. Advances in Cryptology ASIACRYPT 2003 - 9th International Conference on the Theory and Application of Cryptology and Information Security Proceedings. pp. 188-207. New York: Springer.

[10] Baltatu M. & Lioy A. (2001) IP Security. Advanced Security Technologies in Networking. pp. 95-108. Amsterdam: IOS Press.

[11] Federal Information Processing Standards Publication 46-3.

[12] Federal Information Processing Standards Publication 197.

[13] Smith R.E. (2002) Authentication: From Password to Public Keys. Boston, Addition Wesley.

[14] Federal Information Processing Standards Publication 180-1.

ON THE CONDITION MONITORING OF WORM GEARS

Pekka Vähäoja, Sulo Lahdelma and Jaakko Leinonen

University of Oulu, Department of Mechanical Engineering, P. O. Box 4200, FI-90014 Oulu, Finland

Abstract: There has been fairly little research into the condition monitoring of worm gears, although gears of this type have been attached to critical machines in various fields of industry and significant production losses may arise from their breakdowns. Literature has been published on the designing of worm gears but information is rarely provided on fault diagnostics concerning these gears, especially with regard to real-life industrial applications. Some research has been carried out with laboratory-scale test rigs. There is thus an evident need to develop the condition monitoring of worm gears as a basic form of research, but especially to meet the requirements of different fields of industry. This paper gathers together the general information and research reports available on worm gears. Experiences gained from the use of oil analysis and vibration analysis in diagnosing the condition of machines with worm gears are presented in the form of case studies carried in industry. Our experiences of industrial machines with worm gears cover machines employed in cable manufacturing and talc production. We have been monitoring pressers in a cable-manufacturing factory by means of vibration analysis and wear metal analysis of oils for several years now. Condition monitoring data for certain talc agglomerating machines in a talc factory have been gathered during the past few years by means of an oil analysis, and some vibration monitoring data have also been collected. Four similar machines were used here and compared with each other in order to point out similarities and dissimilarities in their behaviour. The aims of this study were to find suitable condition monitoring methods for different machine types all using worm gears and to choose the most sensitive features for routine use in the maintenance of such machines in future. Another aim was to find out if it is possible to propose any common recommendations for the condition monitoring of worm gears.

Key Words: condition monitoring, worm gears, vibration analysis, wear metal analysis

1 INTRODUCTION

A worm gear consists of a worm and a worm wheel, which typically run on crossed axes. The most important part of this gear is the worm, which is usually made of high quality steels. The worm has to be manufactured with high accuracy according to standards. The worm wheel is usually made of different types of bronze mixtures. Due to the high price of bronze, the toothed area is made of bronze and the hub is made of cast iron. By using the worm gear, large gear ratios, even 100:1, can be achieved with only one stage so the relative size and weight of the worm gearbox is usually small. For comparison, helical and conical gears usually have a maximum gear ratio of 5 at one stage in modern applications [1]. Increasing of the gear ratio naturally reduces the diameter of the input shaft, which has a significant influence on the efficiency of the gear. The efficiency of worm gears is usually about 90 %, but it can be as low as 50 % with certain types of worm gears [2]. On the other hand, worm gears damp vibration significantly better than do other types of gears, and run quietly. Worm gears generate a lot of heat, so they have to be cooled, for instance, with a lubricant, ribs of the gear casing or an extra fan. The temperature of the worm gear may not exceed 80 - 90 ^{o}C. Lubrication can be carried out with grease in the case of small and low-speed gears (sliding speed below 4 m/s). Bath lubrication can also be used with low-speed gears (sliding speed below 4 m/s), but splash lubrication is the most common way to lubricate worm gears. Splash lubrication can be used with the sliding speed range from 1.5 to 10 m/s. If the sliding speed exceeds 10 m/s, lubrication has to be carried out under pressure. Lubricating oils for worm gears are usually high-quality, high-viscosity oils with special additives, like EP-additives, to protect or smooth over tooth profiles. Worm gears are often used in machines in which the loading is variable or impact-like. Typical application areas are machines at sawmills and different conveyors. [1-3]

Literature on the condition monitoring of worm gears is relatively scarce. The wearing and surface fatigue of worm gears can easily be detected by means of the wear metal analysis of the gear oil. Wearing of the worm wheel releases bronze particles into the gear oil and correspondingly the worm releases steel particles. The worm wheel usually wears easier because it is made of softer material than the worm. Typical wearing mechanisms are pitting and abrasion [4]. Increased copper and iron concentrations in the gear oil are good failure mode indicators of the wearing of worm wheels and worms in cases where the gear oil only lubricates the gear [5]. Vibration analysis is not used very widely in analysing the condition of worm gears. This may be due to the fact that it is relatively laborious to analyse the vibration signals of worm gears. For instance, impacts at gear mesh frequencies and their sidebands are often followed in the condition monitoring of other types of gears [6], but in worm gears these vibrations are not very informative mostly due to the sliding contact of the gear. Peng *et al.* [7-8] have carried out research where they studied an experimental test rig consisting of a worm gearbox driven by an electric motor in controlled operating conditions. Their focus was an integrated approach to fault diagnosis using both wear debris analysis of gear oil and vibration analysis. The focus was not exactly on the condition monitoring of worm gears, but on comparing the possibilities of these methods in failure detection in more general. They did experiments in different conditions: under the lack of proper lubrication, in normal operating condition and with different contaminant particles in the gear oil, finding out that the wear debris analysis is useful in detecting wear rates and wear mechanisms of the gears, while vibration analysis can provide information on the condition of the bearings. Marshek *et al.* [9] performed a qualitative analysis of the failures of plastic worm gears. They observed that the failure modes for plastic gears were usually quite similar to those of metal gears and differed only in terms of dissimilarities in material properties and processing. Ya-xiong *et al.* [10] studied the wearing of worm gears by analysing the surface shape of the worm tooth flanks. Literature on the usual properties and design of the worm gears is available in the form of machine design textbooks and various articles. For instance, geometric and kinematic factors in worm gear design, tooth contact and stress analysis of worm gears have been discussed in various investigations, and different standards pose requirements that are to be followed. [11-19]

In this study, the condition of certain worm gears was studied by means of a gear oil wear metal analysis and machine vibration analysis. The aim was to find the most suitable features in the failure detection of worm gears and possibly to propose general guidelines for the condition monitoring of worm gears.

2 EXPERIMENTAL PART

The studied machines were used in talc agglomeration in the Mondo Minerals Oy talc factory in Sotkamo, Finland and for copper cable manufacturing in Draka NK Cables Ltd cable factory in Oulu, Finland. Talc agglomerating machines had a worm gear with a worm wheel made of zinc bronze and a worm made of steel. The gears were lubricated with synthetic polyalkyleneglycol oils. The machines had earlier been monitored by means of wear metal analysis and other oil analyses [5]. Cable manufacturing machines had a worm gear with a worm wheel made of tin bronze and a worm made of steel. The machines were lubricated with mineral oils and had formerly been studied by means of wear metal analysis and vibration analysis.

The gear oils were subjected to a wear metal analysis carried out by a microwave-assisted acid digestion and ICP-OES measurement procedure. The gear oil samples (0.4 g) were digested with a 10 ml nitric acid and 2 ml hydrogen peroxide mixture using a two-stage digestion with a Mars 5 microwave oven and HP 500 digestion vessels. The digested samples were diluted with ultra pure water. Typical wear metals were determined by means of Thermo Electron IRIS Intrepid II XDL ICP-OES instrument applying a method described in detail in our other article [20].

Machine vibration measurements were carried out using two different measurement chains. Acceleration signals were detected by means of Brüel & Kjaer Type 4370 and 4370 V accelerometers, amplified using a Brüel & Kjaer Nexus Type 2692 conditioning amplifier and digitized with a NI 9162 Hi-Speed USB carrier, and stored on a lap-top computer with a NI LabVIEW 7.1 software. Acceleration signals were also detected with a Wilcoxon 726 accelerometer and recorded with a Casio DA-7 DAT recorder. Machines were also listened to by means of a SKF TMST2 stethoscope and afterwards from recorded acceleration signals. Effective values of vibration velocity were measured by means of a Mitsol K-94 instrument and peak values of jerk and $x^{(4)}$ by means of a Mitsol D-94 instrument. The same values were also calculated using the LabVIEW 7.1 software. In addition, acceleration, jerk and $x^{(4)}$ time domain signals were analyzed and corresponding spectra were calculated and analyzed with the LabVIEW 7.1 software and also with an Ono Sokki CF 5220 FFT analyzer using an analogue differentiator Mitsol DV-971.

3 RESULTS AND DISCUSSION

3.1 Case Study 1: Talc agglomerating machines

Four similar machines were used in talc agglomeration. Two of them had formerly been studied by means of the wear metal analysis of gear oils. Severe wearing of the worm wheel and the worm of the other machine was detected, while the other was in good condition. Worn machine elements were replaced and the use of both the machines was continued without problems. The detection of the failure at an early stage caused savings of about 200,000 € at diminished repair costs and production losses. The next measurements were carried out ten months after the failure detection, at which point the copper and iron concentrations were at a low level indicating that the repair had been successful. More information on the damaged machine (marked as machine B in this article) is provided in our previous article [5]. At this time, it was decided to monitor the condition of the worm gears of the machines by means of different methods. Gear oil samples were taken from all four machines (machines A-D). In addition, vibration measurements were carried out for two machines (machines C and D).

The analysis results for the two failures, indicating the presence of metals, copper and iron in the gear oils of the machines A-D, are given in Table 1. Copper is the main component in the toothed area of the worm wheels and iron in the worms, which directly indicates the wearing of these machine elements as discussed in our previous article [5].

Table 1

Copper and iron concentrations of the gear oils of talc agglomerating machines

Sample	Cu [mg/kg]	Fe [mg/kg]
Machine A	13	< 2
Machine B	4	< 2
Machine C	94	37
Machine D	< 1.5	5

The concentrations of copper indicating the failure of the worm wheel were at low level in the gear oils of machines A and B and even below the detection limit of the ICP-OES instrument in the gear oil of machine D. However, the copper concentration of the gear oil of machine C is large (94 mg/kg) and indicates that the worm wheel is wearing all the time. It should also be noted that the present gear oil had only been used 14 months. Correspondingly, the iron concentration of the gear oil of machine C has also increased (37 mg/kg) indicating that the worm has also worn slightly. The iron concentrations of the other machines were at a very low level. For comparison, the concentrations of copper and iron in the gear oil in the failure detection of machine B in September 2002 were 106 and 188 mg/kg, respectively. The conclusion is that according to the indicator metal analysis, the teeth of the worm wheel and in a lesser degree also the worm of machine C have worn so their condition should be monitored more intensively.

Vibration measurements were carried out for machines C and D of which the worm gear of machine C was wearing and the worm gear of machine D was in good condition, as indicated by the wear metal analysis. Vibration measurement points are shown in the sketch in Fig. 1 and picture of machine C (machine D is similar) in Fig. 2.

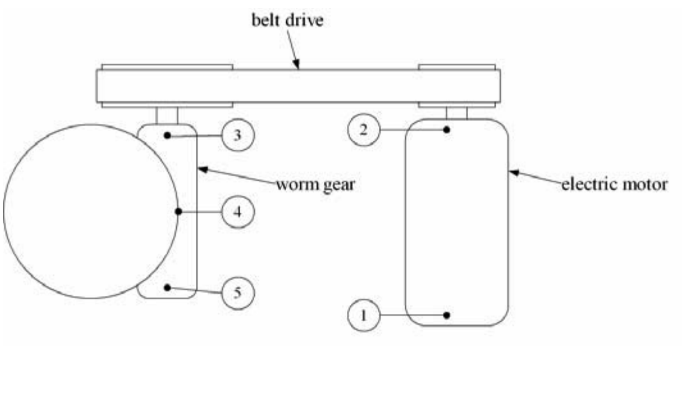


Figure 1. Schematic presentation of machines C and D

Figure 2. Motor (160 kW) and worm gear of machine C.

Excessive vibration at points 3 (vertical), 4 (horizontal) and 5 (vertical) indicates failures in the worm gear, whereas excessive vibration at points 1 (radial) and 2 (radial) indicates the condition of the electric motor. Acceleration signals measured at point 4, i.e. close to the contact point of the worm wheel and the worm, of machines C and D are shown in Fig. 3 and corresponding $x^{(4)}$ signals in Fig 4. More information on the $x^{(4)}$ and other higher order time derivatives of displacement is provided, for instance in [21-22].

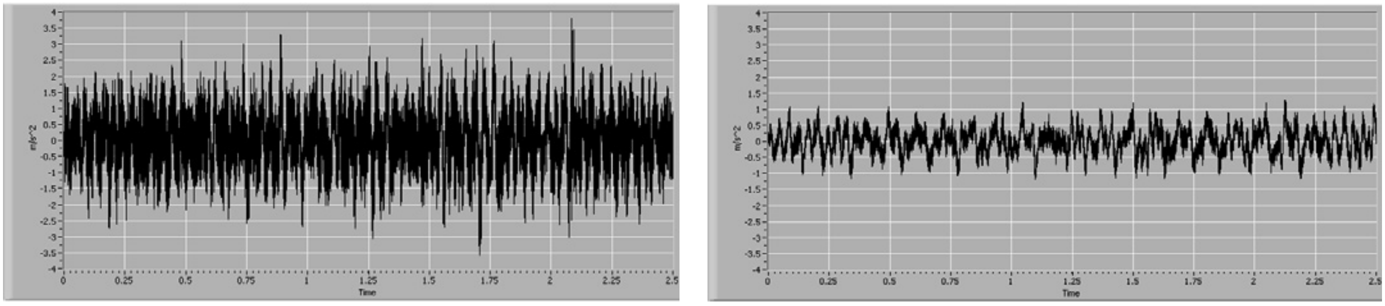


Figure 3. Acceleration signals in the frequency range 3-1000 Hz at point 4 of machines C and D.

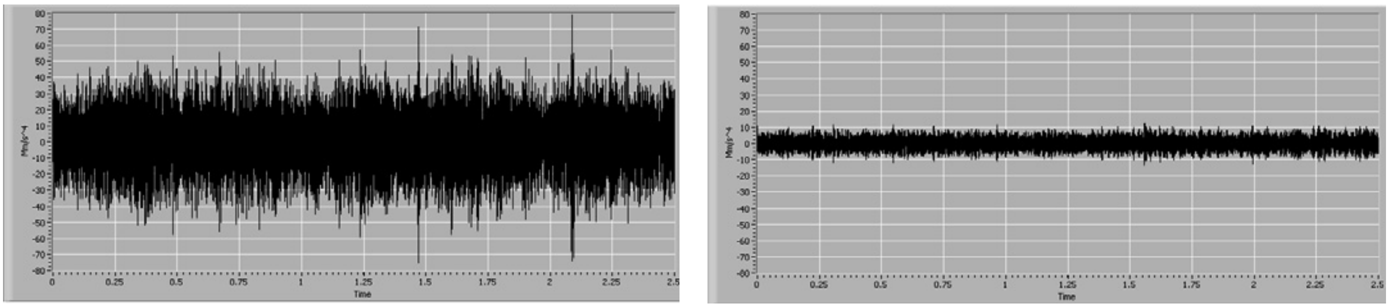


Figure 4. $x^{(4)}$ signals in the frequency range 3-1000 Hz at point 4 of machines C and D.

As indicated in Fig. 3, the levels of acceleration of machine C are larger than the ones of machine D, but there is no clearly distinguishable gear fault. The $x^{(4)}$ signals for machines C and D pointed to a couple of impacts in the signal for machine C. On the other hand, the gear of machine D has no significant impacts. It is easier to detect a gear fault when the frequency range is extended. Similar signals as in Figs 3 and 4 are presented in the frequency range 3-2000 Hz in Figs 5 and 6 and in the frequency range 3-4000 Hz in Figs 7 and 8.

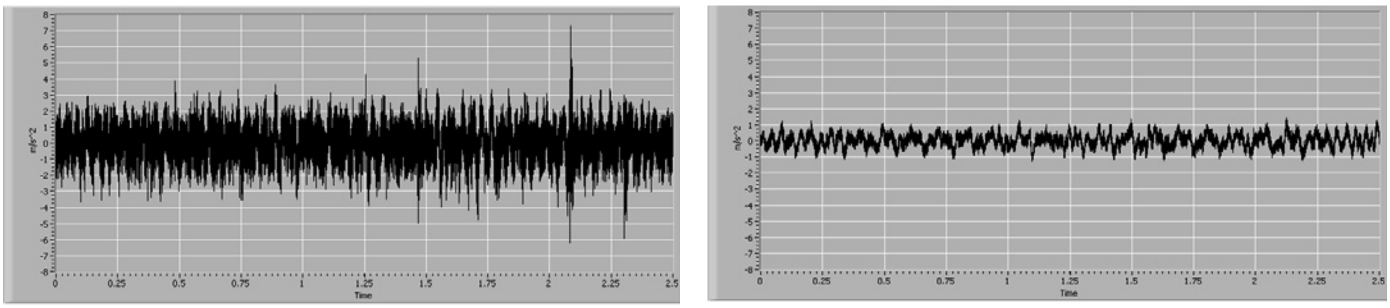


Figure 5. Acceleration signals in the frequency range 3-2000 Hz at point 4 of machines C and D.

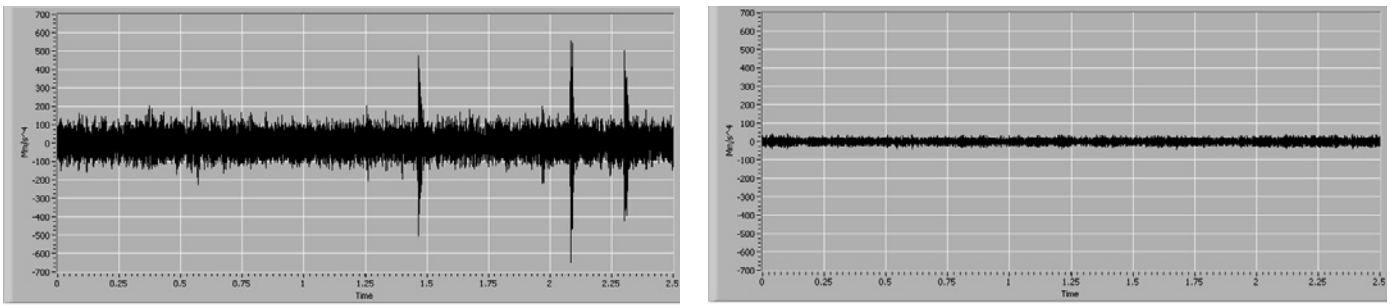


Figure 6. $x^{(4)}$ signals in the frequency range 3-2000 Hz at point 4 of machines C and D.

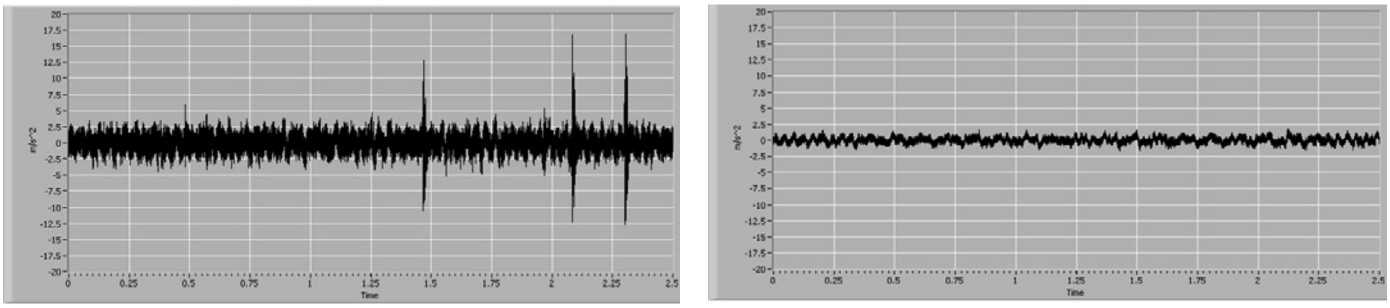


Figure 7. Acceleration signals in the frequency range 3-4000 Hz at point 4 of machines C and D.

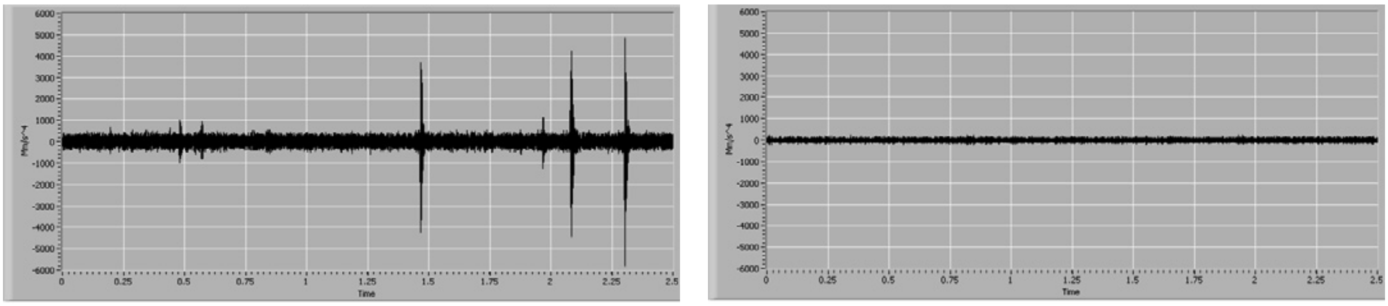


Figure 8. $x^{(4)}$ signals in the frequency range 3-4000 Hz at point 4 of machines C and D.

As seen from Figures 5-6, clear impacts can also be observed from the acceleration signal and very clearly from the $x^{(4)}$ signal in the frequency range 3-2000 Hz. When the frequency range is extended even more, the gear fault can be very easily detected from both the signals. But as stated, it is sufficient to only study the frequency range up to 2000 Hz and the gear fault can still be detected easily. Impacts observed in the time domain signals of machine C could be heard clearly as sharp snaps, when the vibration signals were listened to from DAT tapes and also with a stethoscope. Listening to the signal of machine D only revealed a slight sound of friction, which is typical of worm gears in which the contact between the worm wheel and the worm is sliding. The neighbourhood of the contact point of the worm wheel and the worm, i.e. measurement point 4 in this study, of course tells most about failures and wearing of the worm gear, but the impacts caused by the fault could also be observed easily at the measurement points 3 and 5 on the worm. However, it should be noted that the signals were slightly damped. The reason for the observed impacts is probably the contact of the worm and the failure spots of the worm wheel. The wear metal analysis of the gear oil of machine C already indicated that the worm wheel was wearing, while the gear of machine D had not worn. It was formerly shown in our other publication [5] that the worm wheel of the corresponding talc agglomerating machine (machine B) had worn by pitting, and the pits on the teeth of the worm wheel were large. A finite element analysis made by Su and Qin [23] indicates that certain types of worm gears are sensitive to manufacturing errors. These errors change the contact pattern under loading and increase stress concentration. Simon [24] has indicated that tooth spacing errors and shaft misalignments have influence on tooth contact, so it is possible that slight errors in the tooth profile increase the stress on teeth and finally cause wearing. This wearing can then be observed as excessive vibration in the time domain signals.

One should keep in mind that it seems that the impacts caused by worm gear faults are not necessarily periodical but somewhat stochastic. Because of the behaviour of these gears, it is advantageous to record vibration signals over a relatively long time period. The acceleration and the $x^{(4)}$ signals with a time duration of 50 seconds are shown in Figs 9 and 10.

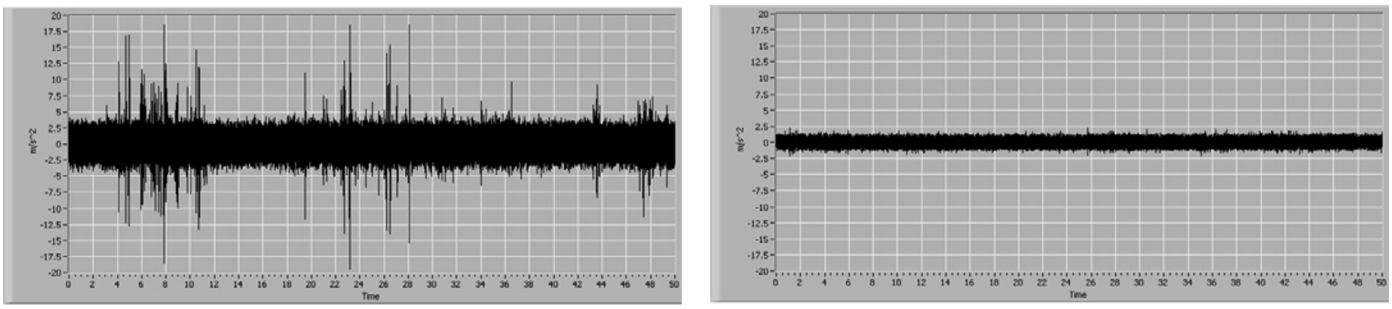


Figure 9. Acceleration signals in the frequency range 3-4000 Hz at point 4 of machines C and D.

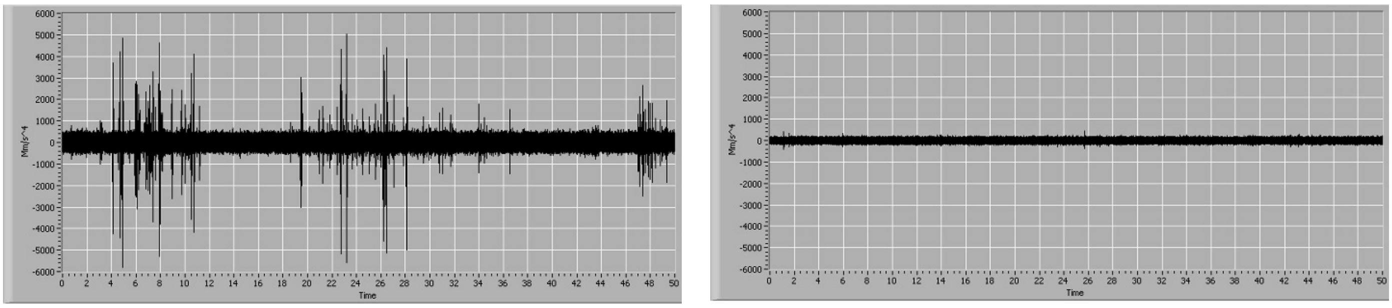


Figure 10. $x^{(4)}$ signals in the frequency range 3-4000 Hz at point 4 of machines C and D.

Figs 9 and 10 show that there is a gear fault in machine C and that machine D is in good condition. The more important observation is that there is no periodic structure in the time domain signals as is usual with faults in other types of gears. The impacts occur almost occasionally. As seen previously, the time domain signals of acceleration and higher order time derivatives of displacement can be very informative, when looking for faults in worm gears. On the other hand, velocity signals were not useful in the failure detection of the worm gears. This is quite obvious, because velocity is seldom sensitive to faults at higher frequencies, like gear failures. The following step after the inspection of the time domain signals is of course to thoroughly study the frequency spectra of the vibration signals. Frequency spectra are usually very informative in detecting failures in other types of gears. If the gear has a fault, increased peaks will occur at the gear mesh frequency, at its multiples and at the sidebands of the gear mesh frequency. Vibration components at the gear mesh frequencies indicate whether or not the gears are in good condition. An example is the acceleration spectrum of an input shaft of a gear in a production crane [25] shown in Fig. 11. All spectra presented in this article are given as rms-values.

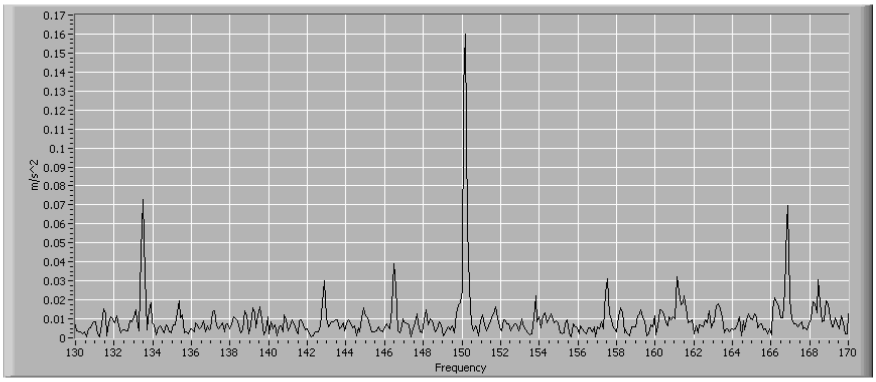


Figure 11. Acceleration spectrum of an input shaft of a gear in a production crane.

An amplitude at the gear mesh frequency of 150.2 Hz and sidebands at 133.5 and 166.9 Hz are observed in this case, although the gear was in good condition. On the other hand, our former study with the copper cable manufacturing machines indicates that a fault in a worm gear cannot necessarily be detected at gear mesh frequencies, but rather as excessive vibration in the higher frequency range caused by increased friction. The frequency range indicating a gear fault was 1350-2250 Hz in that case. Hence it is interesting to see how the frequency spectra can be used in gear fault detection in this case. Acceleration and $x^{(4)}$ spectra of machines C and D are given in Figs 12 and 13, respectively.

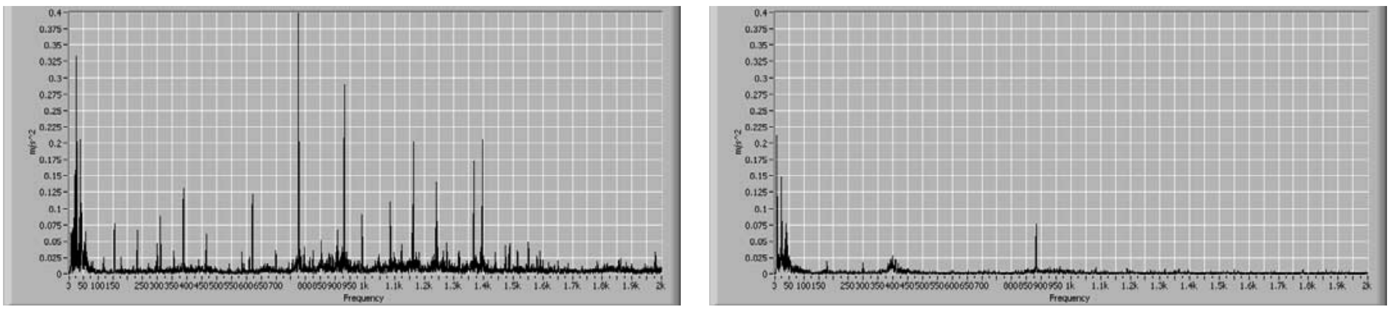


Figure 12. Acceleration spectra in the frequency range 3-2000 Hz of machines C and D at point 4.

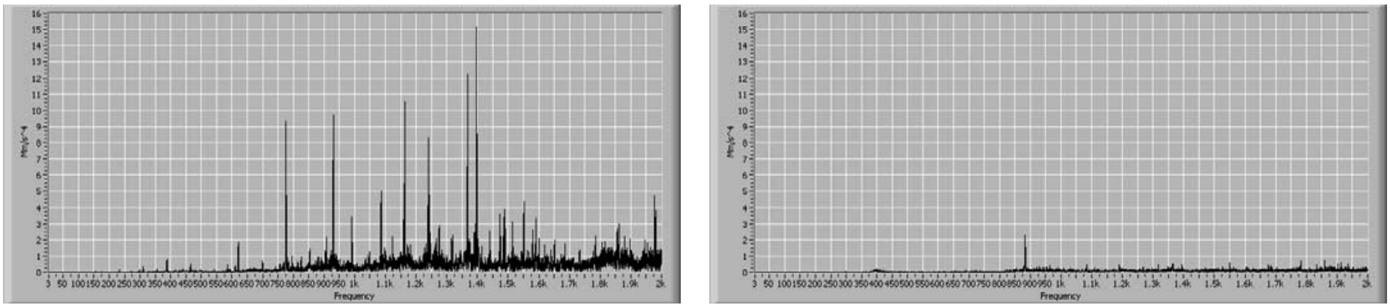


Figure 13. Spectra of $x^{(4)}$ in the frequency range 3-2000 Hz of machines C and D at point 4.

As seen from the frequency spectra of the signals measured near the contact point of the worm wheel and the worm, the amplitudes of acceleration and $x^{(4)}$ are quite low so any exact peak suggesting faults in the gear cannot be observed. Certain peaks at the frequencies of 1000-2000 Hz may be introduced by friction in the worm gear. When the spectrum of either acceleration or $x^{(4)}$ of machine C is compared to the corresponding spectra of machine D, the difference is observable. There are no peaks telling about friction introduced by a fault in the worm gear so the worm gear of machine D is in good condition. The peak at 775 Hz (10 times 77.5 Hz) and other multiples of the frequency 77.5 Hz in the spectrum of machine C suggest an outer race-bearing fault of the electric motor. The vibration caused by this fault is also observed in the worm gear, because the gear and the motor stand on a same foundation *via* which the vibration is transmitted. This fault will be clarified later in this chapter.

One aim of this study was to find out the most sensitive features of worm gear failure detection. Table 2 shows the sensitivities of different features in three different frequency ranges used in failure detection. These sensitivities were calculated from the measurements at point 4. The sensitivities were calculated as averages of twenty values. The sensitivity of a feature is defined as follows:

Sensitivity = (Value of the feature of the machine C) / (Value of the feature of the machine D) (1)

Table 2

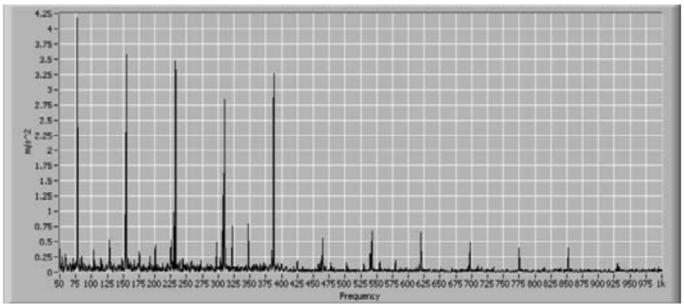
Sensitivities of different features in worm gear failure detection.

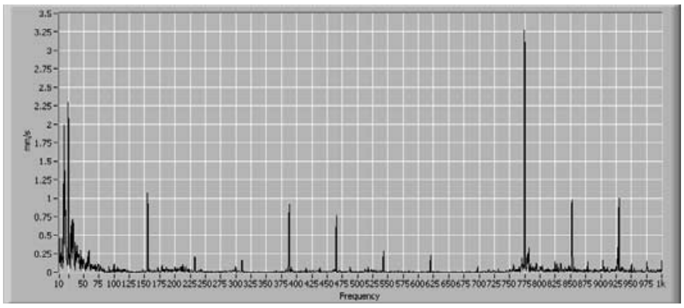
Feature	3-1000 Hz	3-2000 Hz	3-4000 Hz
a_{rms}	2.32	2.77	2.91
$a_{kurtosis}$	1.36	1.27	1.72
a_{peak}	2.93	3.86	5.43
$x^{(3)}_{rms}$	4.18	5.21	3.62
$x^{(3)}_{kurtosis}$	1.14	1.29	4.51
$x^{(3)}_{peak}$	***4.40***	7.06	8.94
$x^{(4)}_{rms}$	4.26	4.98	3.28
$x^{(4)}_{kurtosis}$	1.11	1.65	6.96
$x^{(4)}_{peak}$	4.22	***7.88***	***9.49***

The most sensitive feature in worm gear failure detection in the frequency range 3-1000 Hz was $x^{(3)}_{peak}$, and in the frequency ranges 3-2000 Hz and 3-4000 Hz it was $x^{(4)}_{peak}$. The sensitivity of the peak values in failure detection is very probable, because occasional impacts occur in the vibration signals. The impacts caused by the gear fault are in the higher frequency range and it can also be seen from Table 2 that sensitivities of the features using peak values increase when the frequency range is extended. The rms-values can also be used in failure detection, but they seem to emphasize more the lower frequency range, where e.g. the bearing fault occurs. The sensitivities tend to decrease with $x^{(3)}$ and $x^{(4)}$ using rms-values in the failure detection of the worm gear, when the frequency area is extended from 3-2000 Hz to 3-4000 Hz. Kurtosis seems to be the worst feature in worm gear failure detection in this case. This is in line with our earlier study of worm gears, where kurtosis showed no indication of a worm gear wear [5].

As seen previously in this chapter, wearing of the worm gear could be detected both by means of an analysis of indicator metals (copper and iron) and vibration analysis. Failure in the worm gear might also have been detected by means of other techniques, such as acoustic emission [26]. However, we saw no reason to use acoustic emission in this case, because the failure could be detected with less expensive techniques. It should be noted that the sensors of acoustic emission are more expensive than corresponding acceleration sensors and still vibration analysis should be used to detect faults occurring at the lower frequencies, such as imbalance.

Machine C also had other faults than the wearing of the worm gear. The electric motor has two rolling element bearings and the other of them had an outer race failure and the electric motor also had imbalance. The outer race failure of the bearing could be easily seen from the acceleration envelope spectrum measured at the motor (ball pass frequency of the outer race is 77.5 Hz) given in Fig. 14. Peaks from 77.5 Hz to 10th and even higher multiples can be observed. As a matter of fact, when the acceleration signal was studied by means of an FFT analyzer, 10th and 15th multiples had the highest amplitudes and the amplitudes of these harmonics varied significantly with time. The imbalance of the electric motor was observed at the rotating frequency of 25 Hz. Vibration components occurred also at about 18-19.5 Hz, which is close to the 2nd multiple of the fundamental train frequency of the faulty bearing. The amplitudes of these two frequencies varied randomly when studied by means of an FFT analyzer. This can also explain the variation observed in the effective values of vibration velocity. The velocity spectrum measured at the motor is shown in Fig. 15.





Figures 14 and 15. Acceleration envelope spectrum and velocity spectrum at point 2 of machine C.

3.2 Copper cable manufacturing machines

These two machines had formerly been studied by means of wear metal analysis and vibration analysis [5]. The worm wheel of machine E had worn severely. Both types of analyses detected the failure. More information on this case is given in our other article [5]. The worn worm wheel was replaced and monitoring of the both machines was continued after that. No severe wearing has been detected since and the machines have been working smoothly. An increased amount of copper has been observed in the gear oil of machine E. Before the most recent measurement the wear metal analysis consisted of kerosene dilution of the oil samples and ICP-OES analysis. Now we have developed a new microwave-assisted acid digestion method for degrading oil samples before the ICP-OES analysis. Kerosene does not dissolve solid particles larger than about 10 µm so it is possible that in the case of severe wearing, the results obtained from the wear metal analysis are too small, because particles larger than 10 µm are detached into oil. This is why the new method was developed. As seen from Table 3, the copper concentration of the both machines was quite large in the most recent analysis. The gear oil sample of machine E included solid bronze particles observable with the naked eye, and if analysed, the copper concentration was as much as 120 mg/kg. On the other hand, if the solid particles were ignored, the copper concentration was 41 mg/kg. Which one then is the "right" way to carry out the wear metal analysis? Difficult to say, but it is probable that the value 41 mg/kg represents the copper concentration circulating in the lubrication system. The gear oil sample has to be taken by means of suction from the oil tank, and solid bronze residues from the bottom of the oil tank are also transported into the oil sample. The bronze residues are partly due to a former wearing process of the worm gear and improper flushing of the oil system. Of course, it is evident that the worm wheels of both the machines are wearing, but the situation is not very bad. In any case, monitoring of these machines by means of the wear metal analysis should be continued.

Table 3

Copper and iron concentrations of the gear oils of machines E and F.

Machine	Cu [mg/kg]	Fe [mg/kg]
E (December 2003)	12	< 2
E (November 2004)	28	2
E (November 2005)	41	4
F (October 2003)	36	21
F (November 2004)	22	2
F (November 2005)	45	4

Vibration signals of machines E and F were measured vertically at the bearing of the worm wheel. Acceleration and $x^{(4)}$ spectra of machines E and F in the frequency range 3-2000 Hz are shown in Figs 16 and 17, respectively.

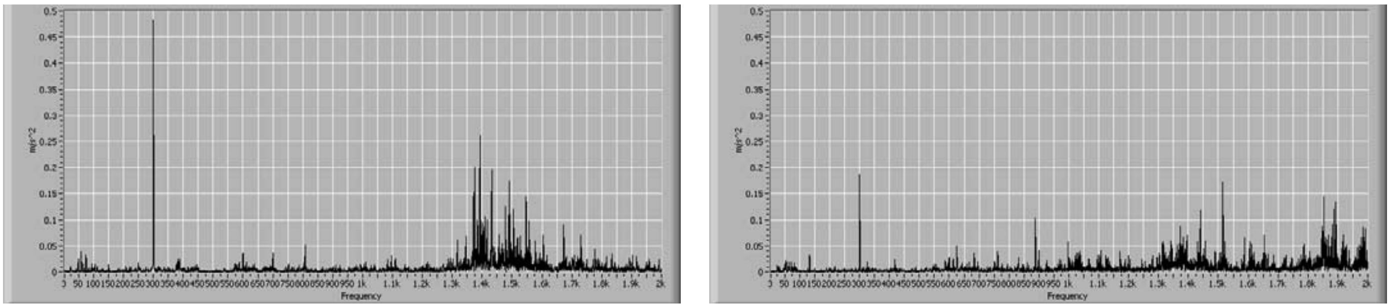


Figure 16. Acceleration spectra in the frequency range 3-2000 Hz at the bearing of the worm wheel of machines E and F.

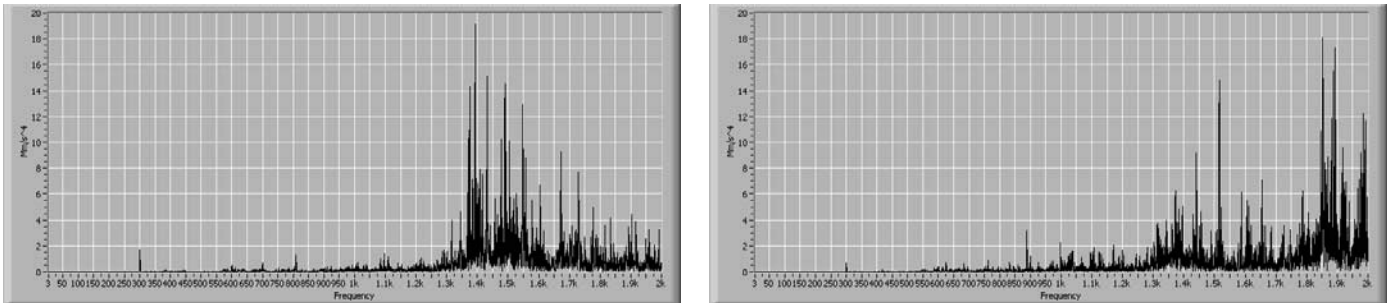


Figure 17. Spectra of $x^{(4)}$ in the frequency range 3-2000 Hz at the bearing of the worm wheel of machines E and F.

When studying the frequency range 1300-2000 Hz in the acceleration and $x^{(4)}$ spectra, numerous peaks can be observed in the spectra of both the machines. When this range was studied by means of an FFT analyzer, significant variation in the amplitudes was observed. This could be an indication of increased friction in the gear. This is in line with our previous study of these machines, in which the wearing of the worm wheel caused vibration in the frequency range 1350-2250 Hz [5]. It is probable that the wearing of the worm wheels, which was observed in the wear metal analysis, will increase friction. The situation may be slightly worse with machine F according to analysing and listening to the signals. A peak at 300 Hz indicates the existence of an electrical fault in the motor of both the machines. The vibration is transmitted in the structure of the machine and seen also at the bearing of the worm wheel. This fault naturally shows well when the measurements are made at the electric motor. The electrical fault is worse with machine E. The v_{rms} measurements in the range 10-1000 Hz do not indicate either the wearing of the worm wheel (the largest values at the output shaft were 0.7 mm/s) or the electrical fault (the largest values at the electric motor were 1.1 mm/s). The most sensitive feature of worm gear failure detection in this case was determined by comparing machines E and F, for which the results are given in Table 4. These sensitivities as averages of twenty values were calculated from the measurements at the bearing of the worm wheel. The sensitivity is defined as follows:

Sensitivity = (Value of the feature of machine F) / (Value of the feature of machine E) (2)

Table 4

Sensitivities of different features in worm gear failure detection.

Feature	3-1000 Hz	3-2000 Hz	3-4000 Hz
a_{rms}	0.58	0.87	1.23
$a_{kurtosis}$	***1.40***	1.04	1.00
a_{peak}	0.77	0.91	1.26
$x^{(3)}_{rms}$	0.98	1.10	1.41
$x^{(3)}_{kurtosis}$	1.12	1.02	1.00
$x^{(3)}_{peak}$	1.08	1.14	1.43
$x^{(4)}_{rms}$	1.25	1.30	1.53
$x^{(4)}_{kurtosis}$	1.16	1.01	1.00
$x^{(4)}_{peak}$	1.33	***1.32***	***1.54***

The most sensitive feature of worm gear failure detection was $x^{(4)}_{peak}$ in the frequency ranges 3-2000 Hz and 3-4000 Hz. This observation is in line with the sensitivities of the same features in detecting worm gear failure in talc agglomerating machines. In the frequency range 3-1000 Hz the most sensitive feature was $a_{kurtosis}$, whereas with the talc agglomerating machines it was $x^{(3)}_{peak}$. The large amplitude at the acceleration spectrum at 300 Hz explains the sensitivity of acceleration in the frequency range 3-1000 Hz. The sensitivity of kurtosis in the frequency range 3-1000 Hz is due to the signal of machine E, which is almost like sine wave having a small value of kurtosis. On the other hand, the signal of machine F has more peaks and therefore a higher value of kurtosis.

4 CONCLUSIONS

Worm gear faults can be identified through the analysis of metals in gear oils and by means of a vibration analysis. Wear metal analysis has been successfully used in detecting the wearing of worm gears in our earlier publication [5]. It is sufficient to measure the concentrations of copper and iron only in order to assure the condition of the worm gear. We have earlier also used vibration analyses in the condition monitoring of worm gears. It is not very easy, though not impossible, to analyse vibration in worm gears The problem is that the impacts caused by faults in the gear do not occur periodically as is usual with other types of gears. Instead they occur almost occasionally so it is beneficial to record the vibration signal over a sufficiently long time period. Measurements should preferably be carried out near the contact point of the worm wheel and the worm, and if this is not possible, at the bearings of the worm or the worm wheel. Vibrations caused by a gear fault could also be observed when measured at the bearings of the worm, but slightly damped in this case. According to our results, worm gear failures can be observed in the frequency range 3-2000 Hz even with acceleration, and the most sensitive one of the studied features of failure detection in this frequency range is $x^{(4)}$ peak value. The rms-values also indicated the gear faults even with acceleration. With the cable manufacturing machines the kurtosis of acceleration was the most sensitive feature in the frequency range 3-1000 Hz, although kurtosis values were not very sensitive with the talc agglomerating machines. When measurements are also carried out at the electric motor, the proper functioning of the whole machine can be assured.

Machines A, B and D at the Mondo Minerals Oy talc factory were in good condition and no wearing of the worm gears was observed. On the other hand, the worm wheel of machine C had worn according to a wear metal analysis. Slight wearing of the worm was also observed. The acceleration and $x^{(4)}$ signals of machine C, which were measured near the contact point of the worm wheel and the worm, revealed intensive impacts that could also be heard when the machine was listened to by means of a stethoscope. These impacts were probably caused by the worm moving over the pits on the worm wheel. In the frequency domain, peaks were observed in 1000-2000 Hz, which could be caused by friction. The electric motor of machine C also had imbalance and a fault on the outer race of one bearing. The bearing had already been replaced and the wearing of the worm gear should be monitored intensively, and the worm wheel should possibly be replaced in future.

The worm wheels of machines E and F in the Draka NK Cables cable factory were wearing, as indicated by the wear metal analysis. Vibration analysis of the machines revealed peaks caused by friction in the frequency range 1300-2000 Hz. The wearing was slightly more prominent with machine F, though both the machines can still be used provided that they are monitored. The electric motors of the both machines had an electrical fault, which was more severe in machine E.

In conclusion, we recommend that worm gears could be monitored as follows.

1. Wear metal analysis and vibration analysis should be combined. The vibration analysis can tell about a fault in the worm gear while the wear metal analysis indicates the wearing machine element.
2. When worm gears are examined, it is enough to only measure the concentrations of copper and iron in gear oil.
3. Vibration measurements should be carried out as near the contact point of the worm wheel and worm as possible.
4. The signal should be recorded over a sufficiently long time period.
5. Time domain signals should be studied carefully and should also be listened to.

5 REFERENCES

[1] Airila M *et al.* (1997) Koneenosien suunnittelu. Porvoo, Finland: WSOY. pp. 546-562. In Finnish.

[2] Dudley DW (Ed.). (1962) Gear Handbook. New York, USA: McGraw-Hill Book Company.

[3] Niemann G. (1965) Maschinenelemente. Berlin, Germany: Springer-Verlag. pp. 151-161. In German. 4 Dudley DW. (1980) Wear in worm gears. In MB Peterson & WO Winer. (Eds) *Wear Control Handbook*. pp. 821-825. New York, USA: ASME.

[4] Vähäoja P, Lahdelma S & Kuokkanen T. (2004) Condition monitoring of gearboxes using laboratory-scale oil analysis. In RBKN Rao, BE Jones & RI Grosvenor. (Eds) *COMADEM 2004*, Cambridge, UK. pp. 104-114. Birmingham, UK: COMADEM International.

[5] Howard PL & Nickerson GW. (1996) Gearbox diagnostic technology. In BKN Rao. (Ed.) *Handbook of Condition Monitoring.* pp. 81-95. Oxford, UK: Elsevier Science Ltd.

[6] Peng Z & Kessisoglou N. (2003) An integrated approach to fault diagnosis of machinery using wear debris and vibration analysis. *Wear*, 255, 1221-1232.

[7] Peng Z, Kessisoglou NJ & Cox M. (2005) A study of the effect of contaminant particles in lubricants using wear debris and vibration condition monitoring techniques. *Wear*, 258, 1651-1662.

[8] Marshek KM & Chan PKC. (1981) Qualitative analysis of plastic worm and worm gear failures. *Wear*, 66, 261-271.

[9] Ya-xiong Z, Ke-jian M, Yan L, Lin Q, Zhan-xiu L & Wei-min Z. (1988) An analysis of the wear features of worm gear tooth flanks. *Tribology International*, 21, 281-285.

[10] Sharif KJ, Kong S, Evans HP & Snidle RW. (2001) Contact and elastohydrodynamic analysis of worm gears. Part 1: theoretical formulation. *Proceedings of the Institution of Mechanical Engineers*, 215, Part C, 817-830.

[11] Sharif KJ, Kong S, Evans HP & Snidle RW. (2001) Contact and elastohydrodynamic analysis of worm gears. Part 2: results. *Proceedings of the Institution of Mechanical Engineers*, 215, Part C, 831-846.

[12] Wang S, Zhan D, Liu H & Wang S. (2002) Tooth contact analysis of toroidal involute worm mating with involute helical gear. *Mechanism and Machine Theory*, 37, 685-691.

[13] Litvin FL, Nava A, Fan Q & Fuentes A. (2002) New geometry of face worm gear drives with conical and cylindrical worms: generation, simulation of meshing and stress analysis. *Computer Methods in Applied Mechanical Engineering*, 191, 3035-3054.

[14] ISO 1122-2:1999. Vocabulary of gear terms. Part 2: Definitions related to worm gear geometry.

[15] ISO/TR 10828:1997. Worm gears – Geometry of worm profiles.

[16] ANSI/AGMA 6034-A87. Practice for enclosed cylindrical worm gear speed reducers and gear motors.

[17] ANSI/AGMA 6030-C87. Design of industrial double-enveloping worm gears.

[18] ANSI/AGMA 6022-C93. Design manual for cylindrical worm gearing.

[19] Vähäoja P, Roppola K, Välimäki I, Kuokkanen T & Lahdelma S. (2006) Development of a Digestion Method and ICP-OES Technique for Wear, Contaminant and Additive Element Analysis of Machinery Oils. *COMADEM 2006*, Lulea, Sweden. Submitted for publication.

[20] Lahdelma S. (1995) On the higher order derivatives in the laws of motion and their application to an active force generator and to condition monitoring. Doctoral thesis. Department of Mechanical Engineering. Research report No 101. University of Oulu. Oulu, Finland.

[21] Lahdelma S, Strackeljan J & Behr D. (1999) Combination of Higher Order Derivatives and a Fuzzy Classifier as a New Approach for Monitoring Rotating Machinery. In J McIntyre & D Sleeman. (Eds) *COMADEM 1999*, Sunderland, UK. pp. 231-241. Oxford, UK: Coxmoor Publishing Company.

[22] Su D & Qin D. (2003) Integration of numerical analysis, virtual simulation and finite element analysis for the optimum design of worm gearing. *Journal of Materials Processing Technology*, 138, 429-435.

[23] Simon VV. (2005) Influence of tooth errors and shaft misalignments on loaded tooth contact in cylindrical worm gears. *Mechanism and Machine Theory*. Article in press. Available on-line at www.sciencedirect.com. Site visited 11.1.2006.

[24] Leinonen J. (2005) Condition Monitoring of Bridge Crane's Hoisting Machinery. Master's thesis. University of Oulu. Oulu, Finland. In Finnish.

[25] Mba D & Rao RBKN. (2005) Development of acoustic emission technology for condition monitoring and diagnosis of rotating machines: bearings, pumps, gearboxes, engines and rotating structures. *The Shock and Vibration Digest*, 37(5), 1-57.

Acknowledgements

The authors are indebted to ABB Oy Service, Draka NK Cables Ltd and Mondo Minerals Oy for their co-operation and to Suomen Ympäristöpalvelu Oy for the possibility to carry out wear metal analyses. Special thanks are due to Ilkka Välimäki, M.Sc.(Chem.), for carrying out the wear metal analyses.

CRITICAL SUCCESS FACTORS FOR THE IMPLEMENTATION OF BUSINESS INTELLIGENCE SYSTEM IN ENGINEERING ASSET MANAGEMENT ORGANISATIONS

William Yeoh[*], Andy Koronios[*], Jing Gao[*]

[*] *Strategic Information Management Lab, School of Computer and Information Science, University of South Australia, Mawson Lakes 5095, Australia.*

Abstract: Much IS literature suggests that various factors play pivotal roles in the implementation of an information system; however, there has been little empirical research about the factors impacting the implementation of business intelligence (BI) systems, particularly in engineering asset management organizations (EAMOs). There is an imperative for a critical success factors (CSFs) approach to enable BI stakeholders to focus on the key issues that leading to successful BI systems implementation. The authors utilised the Delphi method to conduct two rounds of surveys with ten BI system experts of EAMOs domain. Based on the findings, this study identifies ten CSFs that are crucial for implementing BI systems in EAMOs. The paper presents a description and discussion of the CSFs and puts forward recommendations for further research. This study can be valuable to researchers and practitioners who are studying, providing consultancies, planning or implementing BI systems within EAMOs setting.

Key Words: Critical Success Factors, Business Intelligence System, Delphi Study

1 INTRODUCTION

Due to competition resulting from deregulation and from increasingly complex business environments, engineering asset management organisations (EAMOs) such as electricity, gas, water and waste utilities, telecommunications and railway companies are particularly concerned to access timely and high quality information which will enhance decision support and improve the bottom line. Although the computerized transactional systems do provide information to run the daily operations, what EAMOs business executives often need are different kinds of information that can be readily used to make strategic decisions. They need to understand their respective businesses from multiple perspectives, to carry out 'what-if' predictive analysis and to make long-term decisions based on that strategic information. They demand for an easy-to-use intelligent platform and application that is capable of providing strategic information derived from various operational, transactional and legacy systems, spread across multiple platforms and diverse structures. However, these disparate and fragmented operational systems were never designed to support the decision-making process, and hence fail to provide strategic information to the management level. Furthermore, it is impossible to optimise asset lifecycle profits if the management is not provided with a focused approach on asset performance and its interrelationships with other business processes and systems [1]. Yet, with the many millions of dollars of investment in ERP-style systems, EAMOs have been storing vast volumes of transactional data, leading to increased difficulties in analysing, summarising and extracting reliable information. Consequently, in order to solve these problems, EAMOs executives sought to leverage decades-long investments in information systems and data maintenance to improve business decision making. Many progressive EAMOs have turned to implementing BI systems to better manage assets, and this has included transmission and distribution facilities, machines and equipments, supplier and customer relations, and personnel.

A BI system, as Negash [2] explains, is "a strategic information system capable of providing strategic information through a centralised data warehouse, sourced from numerous sources, transformed into meaningful information via analytical and mining tools, to facilitate business insights leading to informed decisions". BI systems come as standardised software packages from such vendors as Business Objects, Cognos, Hyperion Solutions, Information Builders, Oracle and SAS, and they allow customers to adapt them to their specific requirements. A successful BI system implementation presents EAMO with a unified, new and thorough insight across the entire range of its engineering asset management functions. The asset operation management, maintenance management, human resource, inventory control, financial and other business functions are analysed collectively to generate more comprehensive information for better decision support. In other words, critical information from many different sources can be integrated into a coherent body for strategic planning and effective allocation of assets and resources. IDC Research expects that the BI market will enjoy a compound annual growth rate of 23 percent, and report US$3.3 billion dollars turnover in the Asia Pacific region alone [3]. Gartner Research [4] further forecasts that the BI market will continue to grow and flourish through to 2009. This finding is echoed by Darrow's [5] prediction that the worldwide market will reach US$12 billion by 2006.

Despite the increasing interest in, and importance of, BI systems, academic publications relating to this issue are very limited, and this is reflected in the contributions to international conferences and journals. As a BI system is primarily driven by the IT industry, most of the existing BI system literature consists of anecdotal reports or quotations based on hearsay. According to Ang and Teo [6], the implementation of an enterprise-wide information system is a major event and is likely to cause organizational perturbations. This is even more so in the case of BI system implementation. The whole BI project planning and implementation always involves a vast amount of resources and a variety of stakeholders over a period of months, requiring simultaneous attention to a wide variety of organisational, human, budget, technical and cultural variables. Therefore, rigorous academic research to identify the critical success factors for the implementation of BI systems is of great importance. The remainder of this paper has been structured as follows. The next section explains the CSFs approach in information system research, before providing a brief overview of relevant CSFs research in BI systems. The third section then outlines the research methodology for identifying the CSFs of BI systems implementation in EAMOs. In the fourth section the authors analyse and discuss the findings of the study. Then follows the conclusion, and finally proposals for further research initiatives.

2 LITERATURE REVIEW

2.1 Critical Success Factors: Definition and Concepts

The concept of Critical Success Factors (CSFs) was popularised by Rockart [7] as a mechanism to identify the information needs of CEOs. Since then it has become a popular approach in a number of studies linking business and IT [8]. According to Rockart [7], CSFs are those few critical areas where things must go right for the business to flourish. These limited areas are those in which results, if they are satisfactory, will ensure successful competitive performance for the organization. If results in these areas are not adequate, the organisation's effort for the period will be less than desired [9]. The emphasis here is on 'few' and must go 'right'. As the number of CSF is limited, senior management is able to constantly focus on the CSFs until they are successfully achieved [10].

Moreover, Greene and Loughridge [11] add that the identification of CSFs can help to clarify the nature and amount of resources that must be gathered to permit the project team to concentrate their efforts on meeting priority issues, rather than what the available technologies will allow. Evidently, these key areas of activity would need consistent and careful attention from top management if BI system implementation and thus organisational goals were to be attained. In other words, CSFs are high-level management considerations, as distinct from a detailed set of project deliverable specifications. Hence, a set of CSFs identified for the development of any major information system, such as a BI system, is fundamentally different from the set of interlinked detailed tasks which must be accomplished satisfactorily to ensure a project's successful completion [12]. Consequently, the success of BI system implementation may not be fully ensured by meeting the required CSFs, but failure to consider the CSFs will be a major deterrent to success.

Although CSFs was initially introduced to determine the information needs of executives, the approach has been extended to cover other areas of information system research. Despite the limited literature on entire BI system implementation issues, there exists a significant body of CSFs literature about the implementation of information systems, data warehouses, ERP systems and software projects. As a BI system is an enterprise-wide implementation effort similar to ERP and data warehouse, the adoption of CSF approach is therefore considered appropriate in this study of BI system.

2.2 Previous Studies of CSFs for Implementation of BI Systems

The academic literature that explicitly reflects on CSFs for successful BI system implementation is not extensive. The field of BI systems is driven by industry; it takes time for a tradition of academic scrutiny and evaluation to develop into a comprehensive literature. Nevertheless, there have been a number of studies which have sought to identify CSFs for the implementation of an information system, such as a data warehouse and an ERP system. Since data warehouse and ERP system are also involving large IT infrastructure, their CSFs studies are largely applicable to BI systems. The literature forms a basis for the empirical research conducted in this study.

The literature has indicated a number of CSFs for successful systems implementation. In terms of the organisational dimension, support from top management is considered as the most crucial success factor [6, 13, 14]. In particular, Ang and Teo [6] claim that successful implementation completely relies on the strong, persistent commitment of top management. This is because management support can overcome political resistance and encourage proactive participation throughout the organization [13]. Hwang et al [14] further assert that strong executive sponsorship is needed to demonstrate commitment to the BI project, and reinforce the vital linkage of a BI system to business strategy. Additionally, Beath [16] has conducted an extensive study on the importance of a "champion" role in an information system implementation. He indicates that the champion, often the project manager, plays a pivotal role in ensuring the success of a system implementation. Also, Holland and Light [21] note that open communication plays a critical role in ensuring successful implementation. Sarker and Lee [22] confirm through their ERP case study, that systems implementation can only be successful if there is open and honest communication among the stakeholders.

Within the project dimension, Herrmann [17] indicates that project planning and standardisation of activities and documents are needed to minimise the difficulties in collaborating with other organizational units, as well as to reduce coordination and management costs. In addition, the project team needs to ensure that a more complete representation of the entire organisation occurs within the data warehousing project [15], as the BI initiative always involves a variety of organisational stakeholders over a lengthy period. This representation of realities not only applies to the requirement engineering phase, but also to the composition of steering committees to ensure a balanced view of key implementation issues. Furthermore, a balanced multifunctional team is strongly regarded as a key factor for many systems and for the successful implementation of the data warehouse. This team should ideally consist of stakeholders with various skills and solid business backgrounds [13, 18]. Parr et al. [19] indicate that the chosen team members must be fully 'empowered' to achieve their highest possible level of performance. Their assertion is echoed by Levin et al [20], who believe that the empowerment of project team is of great importance to assure faster system roll out. Moreover, the formal and systematic training for end users cannot be taken lightly when aiming for successful system implementation [6, 23].

From the technological perspective, Knox [24] emphasises that management must focus on building BI data environments based on a central data warehouse that allows direct writing of BI applications, constraining datamart creation and introducing standards across data repositories. Ang and Teo [6] recommend that enterprises should start small and plan for continued growth when developing a data warehouse, and that it should be linked to a well-defined BI problem. Otherwise, the value and benefits of the BI project will be difficult to quantify and justify. Watson et al's [15] recent survey suggests that a successful BI deployment must consider scalability requirements from the beginning of the project. Moreover, the project teams are urged to pay significant attention to the transformation of legacy systems and data management tasks [25]. In particular, the transformation and integration of data from EAMO's typically disparate systems is deemed to be the most difficult issue of a BI project. According to Wu [26], a formalised selection plan should also be included to choose the most appropriate tools and technology for the BI system. This plan also provides the project team with a means of checking whether or not the technology was in accordance with business needs. Apart from that, Willcocks and Sykes [27] point out the importance of external consultant support for successful system implementation. They maintain that a qualified consultant must possess knowledge in both business processes and IT, and actively participate in the whole implementation process.

3 RESEARCH METHODOLOGY

The majority of previous studies have used well-established research methods, such as questionnaire surveys and case studies, to examine the key factors for their respective information system research. However, the purpose of present research is not to test hypotheses but rather to identify a set of CSFs for implementation of BI systems, particularly within EAMOs. This study seeks to identify the CSFs that are jointly agreed by a panel of BI system experts who possess substantial experience in EAMOs. The Delphi method was thus deemed to be the most appropriate method for this study [28], because it allows the gathering of subjective judgements which are moderated through group consensus in the absence of rich literature on BI systems. To be specific, a Delphi study is a structured group-communication method for soliciting expert opinion about complex issues, such as CSFs for BI systems implementations, through the use of a series of questionnaires and controlled feedback [29]. Moreover, the method is particularly suitable for this research situation where personal contact between panelists is not desirable because of concerns about the difficulty of ensuring democratic participation.

For data collection, a Delphi panel composed of ten BI systems experts in EAMOs domain was established. Ziglio [30] asserts that useful results can be obtained from small homogeneous groups of 10-15 experts, and that the size and composition of such a panel is thus deemed suitably representative. As shown in Table 1, the panelists have been involved substantially in the implementation of BI systems within EAMOs in Australia and the United States. They consist of senior consultants from BI consultancy firms and BI project managers and system managers within implementation companies. The BI systems included Cognos, Business Objects, Oracle, Hyperion, Microstrategy, SPF Plus, Actuate, Informatica and IBM.

In addition, the range of engineering asset management organizations represented in the implementation by these experts was diverse. The EAMOs include public utilities, such as electricity, gas, water, and waste management, and infrastructure-intensive enterprises such as telecommunications and railway companies. It should be noted that some of the large organisations have had several implementation projects and that they chose to implement the BI systems in a series of phases. Most of the EAMOs are very large companies with engineering assets worth hundreds millions of dollars and have committed immense expenditure to BI projects. So the expertise of the panelists represents 'state of the art' knowledge of BI systems implementation in a broad range of engineering asset-intensive industries.

During the first round of the Delphi study, a semi-structured interview session was conducted separately with each panelist. The interview process followed the focused interview format in which the interviewer followed a set of structured questions derived from literature review. However, the existing literature is not a comprehensive research on an entire BI system, but mainly focuses on data warehousing, which is the core component of a BI system. Therefore, the panellists were allowed - and indeed encouraged - to express their insights into other relevant CSFs. After the interview, further clarifications (if any) were made by follow-up phone calls and email communications. However, the authors were cautious about being over-reliant on panellists, and attempted to seek further evidence to corroborate emerging findings and issues from the documentation and presentation materials provided by the panel. Subsequently, the data gathered from first round of interviews was analysed thoroughly. As a result, a set of candidate CSFs was produced by combining all different factors perceived by the panel to be important.

In the second round of the Delphi study, the combined list of candidate CSFs was surveyed among the panel using a structured questionnaire approach. Specifically, the combined candidate CSFs was tested by elicitation of a Boolean scale in the process of seeking consensus from the BI experts. The purpose of using Boolean response (the factor is/is not critical for successful implementation of a BI system in EAMOs) aims to distinguish important factors from critical success factors. At this round, the majority of the panelists (that is, more than 70 percent of the panel members) shared a common opinion on the short listed CSFs that will be discussed in the following section. Sumsion [31] points out that opinion stability reflects consensus when the panel disagreement is less than 30 percent and that this offers a working definition of a threshold for stability, and hence this round of the Delphi study is considered to have reached a critical point for research termination.

Table 1

Delphi Panellists and their BI Systems Experience in EAMOs

Current Position	Current Company Type	BI System	EAMOs' Industry Sector
Principal Consultant	BI Consultancy	Business Objects, Information Builder, Cognos, Oracle	Electricity utility, Gas utility, Water utility Waste water utility, Power generation, Public transportation authority
Principal Consultant	BI Consultancy	Cognos, Business Objects, Actuate	Telecommunications, Municipal utility, Airlines
Senior Consultant	BI Consultancy	Oracle	Electricity generator and distributor
Business Information System Manager	Implementation company	Informatica, Oracle	Water utility, Waste water utility
Senior Principal consultant	BI Consultancy	Cognos, Hyperion, Microstrategy, Business Objects	Telecommunications, Electricity utility, Gas utility, Water utility
Senior Advisor	Implementation company	Cognos, Oracle	Rail infrastructure and fleets
Consultant	BI Consultancy	IBM, Oracle	Municipal utility, Public transportation authority
System Manager	Implementation company	SPF Plus	Electricity utility
Project Manager	Implementation company	Oracle, Hyperion	Rail infrastructure and fleets
System Support Engineer	BI Vendor	Cognos	Energy utilities

4 DISCUSSION OF FINDINGS

This section discusses the critical success factors perceived by the panel of BI experts in the implementation of BI system within EAMOs. In particular, the panellists were directly drawing on considering their hands-on experience in EAMOs' BI system implementations. Although many factors were elicited in the first round Delphi study, only CSFs which received a panel consensus of more than 70 percent are reported here. These CSFs provide important considerations for relevant stakeholders of EAMOs working on BI system and they help them to avoid costly pitfalls. The details are discussed below.

4.1 Top Management Support and Sponsorship

Top management support was held to be indispensable for successful BI system implementation in EAMOs. The panel asserted that consistent support and sponsorship from top management make it easier to secure the necessary operating resources such as funding, human skills, organisational support, and other requirements throughout the implementation process. The whole BI system implementation effort is a costly, time-consuming, resource-intensive initiative. Hence, the presence of sufficient resources can help the project team to overcome implementation obstacles and deliver on the project milestones. In particular, adequate funding would enable the project team to access the appropriate BI tools and technology. The experts indicate that the selection of BI system components such as ETL tools, data-cleansing software, database management systems, data warehouses, query tools, analytical tools and hardware that are needed for optimized BI application must not be jeopardized because of a lack of funding. Otherwise, the project scope will have to be realigned accordingly and this may present a threat to user expectations. Moreover, without dedicated support from senior executives, the BI project may not receive the proper recognition and hence the support it needs to be successful. This is simply because users tend to conform to the expectations of top management, and they are more likely to accept a system backed by their superiors. Many panellists further indicated that such support was best when it was accompanied by close monitoring of, and interaction with, the BI projects. In this regard, a steering committee was established to provide a balanced check between management needs and project effort. Specifically, senior management set up a structured mechanism for regular monitoring of implementation progress. Apart from these, top management support was deemed critical to solve the organisational resistance and conflicts that may arise for such a complex project.

4.2 Balanced Team skill and Composition

Owing to the immense expense, and the steep learning curve involved in implementing a BI system, the skills, knowledge, experience and aptitude of the development team are perceived to have a major influence on the outcome of the project. The panel stressed that the project team should be composed of those personnel who possess technical expertise and those with a strong business background and knowledge, because BI is a business-driven project to provide enhanced managerial decision support. The panel maintained that these project members must desire to work as a collective team and be able to work and interact well with users, demonstrating excellent interpersonal skills and communication capabilities. In addition, many experts indicated that it is much better off empowering the project leader to choose his or her best team. An effective model has been the small team comprising five to eight people who are knowledgeable, skilled, and dedicated, and who like and respect each other in order to strive for project success. Moreover, these team members should be available full time and released from all other assignments to work on the implementation.

4.3 Project planning and scope definition

According to the panel, for successful BI system implementation it is necessary to carefully take into account the important factors of project planning and scope definition. Proper project planning facilitates flexibility and the adaptability to changing requirements within the given timeframe and resources, especially in implementing a complex EAMO-specific BI system. The panel emphasized that planning provides guidelines which enable the project team to focus on crucial milestones and pertinent issues while shielding them from becoming tapped in unnecessary events. Furthermore, the implementation effort typically requires massive changes in other technologies that are linked to it. Hence, it is critical to expect the project team, and particularly technical members, to provide disciplined preparation for, and completion of, technical aspects to aid in the introduction of changes to existing technologies within the planned timetable and budget. Moreover, the experts strongly recommended that all discussions on key issues should be documented and communicated to relevant BI stakeholders so as to avoid misunderstanding and denial of responsibilities. One consultant pointed out that the "Just do it" philosophy and the excuse of "What a waste of time developing a detailed project plan, yet no one follows exactly to the plan" are warning signs of project failure. In addition, many experts further indicate that it is advisable to start small and adopt an incremental change approach, as large-scale change efforts are always fraught with greater risks given the substantial variables to be managed simultaneously.

4.4 Building a pilot system and incremental change

The panel unanimously agreed that a pilot system is always valuable as proof of a concept; that is, constructing a fairly small BI application for a key area in order to provide tangible evidence for both executive sponsors and general users. They perceive a pilot system that offers clear forms of communication and a better understanding in an important business area would convince organizational stakeholders on the usefulness of BI system implementation. As a result of a successful pilot project, senior management are keen and motivated to support larger scale BI efforts. Moreover, a properly-conducted pilot system is not a wastedproject; for it could be further expanded and enhanced into an enterprise-wide BI system. The pilot system could include additional data sources, attributes and dimensional areas for fact-based analysis and could incorporate external data from suppliers, regulatory bodies and industrial benchmarks. Undoubtedly, a small and highly-focused pilot system may allow an organization to concentrate on crucial issues, so enabling teams to prove that the system implementation is feasible and productive for the enterprise.

4.5 Modeling of Dimensional Data and Metadata

A BI system provides EAMOs users with a dimensional view of data such as engineering assets, and each of the dimensions has measures in terms of throughput or productivity. In order to have consistent dimensions and measures across subject areas, the panel asserted that it is important for users to identify and establish consensus on the dimensions and measures that will be used in the system's data model. Prior to that, many experts stressed that the information needs of organizational key users must be identified to enable the decision regarding the data requirements which will satisfy their expectations and information needs. Consequently, the project team would use those requirements to develop an enterprise-wide data model that is business orientated. Furthermore, it is typical for an EAMO to have hundreds of varying terms, with slightly different meanings, as business units tend to define terms in ways that best serve their purposes. For instance, a maintenance unit might define assets as physical engineering equipment or machine, whereas a financial department might include capital and buildings. Therefore, the development of a metadata model on which to base the logical and physical data warehouse construction for BI system will ease terminology problems. The metadata model, according to the experts, must be sustainable for a long period to provide the consistency upon which end-users rely. More importantly, the metadata model should be flexible enough to enable the scalability of the BI system while consistently providing integrity on which OLAP, data mining and executive information systems depend.

4.6 Formal and Interactive User Involvement

The experts also perceive that with better user participation included in the implementation effort, this can lead to better communication of their needs, which in turn can help ensure the system's successful implementation. This is particularly important when the requirements for a system are initially unclear, as is the case with many of the decision-support applications that a BI system is designed to sustain. Significant numbers of panellists explained that user participation can help meet the demands and expectations from various end users, and one panellist asked, "How can the project team design and implement a BI system to meet the users' needs without their involvement?" Hence, the data dimensions, business rules, metadata and data context that are needed by business users must be consulted and incorporated into the system. Another panellist mentioned that "…if the BI project does not involve adequate user participation, then the project is not likely to be successful because the BI application will not satisfy the end users." Besides that, when users are actively involved in the project, they have a better understanding of the potential benefits and this makes them more likely to accept the system on completion.

4.7 Data Quality and Reliable Sources

The Delphi study indicates that the quality of data is crucial if a BI system is to be implemented successfully. This can be anticipated, as data quality will affect the quality of management reports, which in turn impact the decision outcomes. A primary purpose of the BI system is to integrate disparate data in the organisation for further analysis via BI analytical tools. However, most EAMOs face great difficulties if it is found that the data in their source systems are of poor quality. These problems may be due to the lack of concern on the part of field maintenance workers and a lack of enforcement by managers. Thus, the combined efforts of the project team and business units are required to assess the quality of the data and to put appropriate data-cleansing processes in place, especially during the ETL process Furthermore, the EAMO should initiate efforts to improve the quality of the data in such source systems as operational, maintenance, procurement, contract management, work management, finance and human resource, because limitations will have a ripple effect on the BI applications and subsequently the decision outcomes.

4.8 Formal Selection of Development Tools and Technology

The panel further confirmed that the selection of development tools and technology impacts the efficiency and effectiveness of BI development efforts as much as other factors, particularly if the tools are not well understood by the project team. In fact, the tools and technologies involved in BI systems within EAMOs are different from those used with operational systems because the BI effort requires advanced data extraction, transformation and loading software, data-cleansing programs, database software, and multidimensional analysis tools. Many experts assert that the whole implementation effort will suffer if the development technologies do not meet the requirements of strategic asset management for EAMOs. Therefore, the selection of these important development technologies and tools must meet the needs and expectations of asset managers and users. Some consultants suggested that using a formal business-based methodology approach to select against defined goals can significantly increase the success rate of getting most optimised tools and technology. This business-driven approach builds upon consensus and therefore is more likely to meet user requirements. Furthermore, one panellist advised that the project team should not let an external consultant make important decisions on matters such as software selection. The skills and experience of the consultant may hinder the individual's judgment from offering an independent recommendation.

4.9 Formal and Adequate User Training

The panel also agreed that the formal and systematic training for end users must not be ignored when aiming for successful BI system implementation. Many experts indicated that adequate training can help users to speed up the system adoption and to explore the required information in a more effective manner. One expert posed the following question, "Will the users of BI system know how to use it? If not, how are they going to make use of the BI application?" Furthermore, panellists emphasized that training should focus on the technology itself as well as on the associated management and maintenance issues. This training is important to equip users to understand and experience the features and functions and to learn about the configured environment and business rules of the BI applications. In order to facilitate more conducive training sessions, it is also critical to separate power users from casual one because individuals may have different needs and expectations. Moreover, the experts stressed that companies that view training as an investment rather than a cost, reap greater benefits from the project implementation because users are accountable for making the system produce timely and accurate information.

4.10 Presence of a Champion

Apart from the above CSFs, the panel agreed that getting the right person to lead the project is critical for implementation success. The project leader, or so-called 'champion', must not only be technically-knowledgeable, but possess in-depth business acumen as well as excellent interpersonal skills to deal with organisational conflicts. He or she must be committed to the success of the project, must be unswerving in promoting the benefits of the new BI system, and must help resolve any internal disputes that are impacting the project. Also, the champion must be able to motivate his team to channel their energies into overcoming both technical and organisational obstacles to better meet project objectives. Consequently, the selection of a champion may determine the cohesion and success of BI system implementation.

5 CONCLUSION AND FURTHER RESEARCH

This paper has reported the results of a Delphi study that aimed to critically identify the CSFs for the implementation of BI systems in EAMOs. Since this research topic is a relatively new area, the approach drew upon a collection of expert opinions through interviews with 10 BI practitioners who have been involved extensively in implementing BI system within EAMOs. Based on the first round interview findings, the authors then conducted a questionnaire survey with the panel in order to seek consensus on the candidate CSFs. Finally, a list of CSFs was produced from the Delphi survey result that met the minimal 70 percent panel agreement. An analysis of the findings demonstrated that there is a combination of CSFs peculiar to successful BI system implementation in EAMOs. Senior management, project sponsors, project leaders, project members, end-users and external consultants all play critical roles in respective CSF. While the specific CSFs vary somewhat among BI systems and general IS studies, it appears that there is a new understanding of factors associated with infrastructure projects like BI systems which can be used in future research. Many CSFs of BI system implementation are probably the same as the IS development. However, there is a great difference with the technical CSFs like data modeling issue because the technical issues vary with the nature of the infrastructure system.

From a practical standpoint, the findings of this study help BI project stakeholders both to identify and to concentrate on the CSFs, especially in the planning of a BI system. Such outcomes will help them to improve the effectiveness and efficiency of their implementation activities by obtaining a better understanding of possible obstacles that might hinder the successful BI system implementation. It is also believed that by understanding the CSFs for their system the organisation is more likely to have a more reliable BI system after its implementation. For further research, it is planned that a survey with multiple EAMOs will be conducted using a structured questionnaire based on this Delphi result. As it stands, this research was exploratory in nature and further empirical research in other industries may shed light on where and how the research findings can influence the success of BI system implementation.

6 REFERENCES

[1] Eerens E. (2003) Business driven asset management for industrial & infrastructure assets. Victoria, Australia: Le Clochard.

[2] Negash S. (2004) Business intelligence. Communications of the Association for Information Systems, 13, 177-195.

[3] Giang R. (2002) Asia/Pacific Business Intelligence Solutions Market Spending Dynamics. IDC Research Group.

[4] Buytendijk F, Dresner H & Linden A. (2004) Hype Cycle for Business Intelligence. Gartner Research.

[5] Darrow B. (2003, Feb 3) Making the Right Choice - Solution providers are evaluating a plethora of options as they puzzle over the future of business intelligence. Computer Reseller News.

[6] Ang J & Teo. (2000) Management issues in data warehousing: insights from the Housing and Development Board. Decision Support System, 29(1).

[7] Rockart J. (1979) Chief Executives Define Their Own Data Needs. Harvard Business Review, 57(2), 81-93.

[8] Rokart J & Hoffman J. (1992) System delivery: Evolving new strategies, Sloan Management Review, 33(4), 21-31.

[9] Ives B, Jarvenpaa S & Mason R. (1993) Global business drivers: Aligning information technology to global business strategy, IBM System Journal, 32(1), 4-16.

[10] Khandelwal V. (2000) What the Australian CEOs want from IT, Journal of Research and Practice in Information Technology, 32(3), 151-167.

[11] Greene F & Loughridge F. (1996) The Management Information Needs of Academic Heads of Department in

Universities: A Critical Success Factors Approach'. A Report for the British Library Research and Development Department. Sheffield: Department of Information Studies, University of Sheffield.

[12] Dobbins J. (2000) On a Generalized CSF Process Model for Critical Success Factors Identification and Analysis, PhD Thesis, George Washington University.

[13] Wixom B & Watson H. (2001). An empirical investigation of the factors affecting data warehousing success. MIS Quarterly, 25(1), 17-41.

[14] Hwang H, Ku C, Yen D & Cheng C. (2004. Critical factors influencing the adoption of data warehouse technology: a study of the banking industry in Taiwan, Decision Support Systems, 37(1), 1-21.

[15] Little R & Gibson M. (2003) Perceived influences on implementing data warehousing. IEEE Transactions on Software Engineering, 29(4), 290-296.

[16] Beath C. (1991) Supporting the Information Champion. MIS Quarterly, September, 355-371.

[17] Herrmann C. (2004).Exploring the Strucutural Dimension of Data Warehouse Organizations: Results of a Survey and Implications. Proceedings of The IFIP TC8/WG8.3 International Conference, Tuscany.

[18] Lee Y, Lee D & Suh C. (2000) Factors affecting implementation of data warehouse, Proceedings of The Pacific Asia Conference on Information System, Hong Kong.

[19] Parr A, Shanks G & Darke P. (1999) Identification of Necessary Factors for Successful Implementation of ERP Systems. New Information Technologies in Organizational Processes: Field Studies and Theoretical Reflections on the Future of Work, Kluwer academic publishers, chapter 8, 99-119.

[20] Levin R, Mateyaschuk J & Stein T. (1998) Faster ERP Rollouts. Information Week, July.

[21] Holland C & Light B. (1999) A framework for understanding success and failure in Enterprise Resource Planning System Implementation. The Proceedings of 7th European Conference on Information Systems, Copenhagen.

[22] Sarker S & Lee A. (2000) Using a case study to test the role of three key social enabales in ERP implementation. Proceedings of the 21st International Conference on Information Systems, Brisbane.

[23] Joshi K & Curtis M. (1999) Issues in the planning of a data warehouse. Information Strategy, 15(2), 28-35.

[24] Knox, M (2004) Asset Managers are Building Their BI Environments, Gartner Research, April.

[25] IBM (2003) Transforming Data into Business Intelligence - An Overview, Presentation for State of Louisiana Department of Transportation, viewed 28 September 2005, <www.lacisd.org/ppt/tech-day-ibm-trans-data-Jun03.ppt>.

[26] Wu J. (2000) Business Intelligence: Top 10 Warning Signs of a Troubled BI Project,

[27] viewed 20 September 2005, <www.dmreview.com/editorial/dmreview/print_action.cfm?articleId=2707>

[28] Willcocks L & Sykes R. (2000) The Role of the CIO and IT Function in ERP. Communications of the ACM, 43(4).

[29] Linstone H & Turoff M. (2002) The Delphi Method: Techniques and Applications, Reading, M.A: Addison-Wesley Publishing Company Inc.

[30] Day J & Bobeva M. (2005) A Generic Toolkit for the Successful Management of Delphi Studie, The Electronic Journal of Business Research Methodology 3(2), 103-116.

[31] Ziglio E. (1996) The Delphi Method and its Contribution to Decision-Making. in: Alder M.& Ziglio E. (Eds.) Gazing into the Oracle, The Delphi Method and Its Application to Social Policy and Public Health. London: Jessica Kingsley Publishers, 3-33.

[32] Sumsion T. (1998) The Delphi technique: an adaptive research tool. British Journal of Occupational Therapy 61(4), 153-156.

Acknowledgments

This research is conducted through the Cooperative Research Centre for Integrated Engineering Assets Management (CIEAM). The support of CIEAM partners is gratefully acknowledged.

FEATURE GROUP OPTIMIZATION FOR MACHINERY FAULT DIAGNOSIS BASED ON FUZZY MEASURES

Xiaofeng Liu, Lin Ma, Sheng Zhang and Joseph Mathew

CRC for Integrated Engineering Asset Management, School of Engineering Systems,

Queensland University of Technology, Brisbane, QLD 4001, Australia

Abstract: With the development of modern multi-sensor based data acquisition technology often used with advanced signal processing techniques, more and more features are being extracted for the purposes of fault diagnostics and prognostics of machinery integrity. Applying multiple features can enhance the condition monitoring capability and improve the fault diagnosis accuracy. However, an excessive number of features also increases the complexity of the data analysis task and often increases the time associated with the analysis process. A method of bringing some efficiency into this process is to choose the most sensitive feature subset instead. Fuzzy measures are helpful in this regard and have the ability to represent the importance and interactions among different criteria. Based on fuzzy measure theory, a novel feature selection approach for machinery fault diagnosis is proposed in this paper. A heuristic least mean square algorithm is adopted to identify the fuzzy measures using training data set. Shapley values with respect to the fuzzy measures are applied as importance indexes to help choose the most sensitive features from a set of features. Interaction indexes with respect to the fuzzy measures are then employed to remove the redundant features. Vibration signals from rolling element bearing test rig are used to validate the method. The results show that the proposed feature selection approach based on fuzzy measures is effective for fault diagnosis.

Key Word: Fault diagnosis, Feature selection, Fuzzy measures, Importance index; Interaction index

1 INTRODUCTION

The task of condition monitoring and fault diagnosis for modern machinery is becoming increasingly complex as the machines become more automated and complicated. Fortunately, with the development of modern multi-sensor technology and advanced signal processing techniques, the increasing complexity of the tasks of diagnostics and prognostics of these machines can now be managed more meaningfully.

The task of condition monitoring and fault diagnosis usually results in a probabilistic result with a level of uncertainty. According to information fusion theory, information uncertainty can be reduced by applying multi-source information [1]. Hence, condition monitoring and fault diagnosis capability can be enhanced by applying multiple features. However, an excessive number of features on the other hand also increases the difficulties of data analysis and usually escalates maintenance cost. In fact, using all the features for monitoring and diagnosis purposes is also unnecessary as the contribution of each feature varies in a range. Some of them may be of little use to a specific troubleshooting process. Thus, selecting the most valuable feature subset is extremely important.

Various methods have been applied for feature selection purposes. Sub-optimal methods are more widely used than optimal methods. Although sub-optimal methods can generally produce optimal or near-optimal results in most real application cases, they cannot guarantee optimal results. The most frequently used sub-optimal feature selection algorithms are sequential forward search, sequential backward search, plus-L-minus-R and floating search, etc [2].

The exhaustive search is a universal optimal method. It produces an optimal solution by testing all the possible feature combinations. However, as the number of feature combinations increases exponentially with the number of features, this method becomes difficult to implement for high dimension feature space [3] [4] [5]. The only method other than exhaustive search still capable of yielding an optimal result is the branch and bound (BB) based algorithm. However, its application is also conditional. This method is only limited to monotonic criteria. Furthermore, its converging speed is highly dependent on the data sets [3]. To enhance its performance, BB algorithms have been improved by some researchers [6, 7].

Classical set theory and probability theory have been widely applied for dealing with uncertainty problems. Fuzzy set theory and fuzzy measure theory are two more general mathematical methods. Fuzzy methods are more effective than traditional clustering methods in handling fault features which are imprecise, with the boundaries among different failure modes usually being ambiguous in their mapping space. Fuzzy methods describe fault patterns in a non-dichotomous way which is similar to the manner in which human beings process vague information. As an outgrowth of classical measure theory, Fuzzy Measure (FM) and Fuzzy Integral (FI) theory has been applied to pattern recognition [8] [9] [10], image processing [11] [12] [13] and information fusion [14] and have the advantage that they are able to represent importance of criteria and certain interactions among them.

This paper presents a novel optimal feature selection approach for machinery fault diagnosis based on the fuzzy measure theory. It is suitable for low feature space problems. A heuristic least mean square algorithm (HLMS) is adopted to identify the fuzzy measures from training data set. Shapley values with respect to the fuzzy measures are computed and applied as importance indexes to choose the most sensitive features from a number of features. Interaction indexes with respect to the fuzzy measures are then employed to remove the redundant features. Vibration signals from rolling element bearings are used to validate this method. The results show that the proposed feature selection approach based on fuzzy measures is effective for fault diagnosis and is agreeable with feature subset obtained through other method.

The rest part of this paper is organised as follows. In section 2, the definition of fuzzy measures and related concepts, such as Shapley value and interaction index, are briefly introduced. In Section 3 a heuristic least mean square algorithm for identifying 2-additive fuzzy measures is introduced. A novel approach for fault diagnosis feature optimisation based on fuzzy measures is proposed in section 4. Section 5 presents the results of the experimental evaluation. Discussions based on the results are presented in Section 6, and Section 7 draws the conclusion.

2 FUZZY MEASURE, CHOQUET FUZZY INTEGRAL, IMPORTANCE INDEX AND INTERACTION INDEX

A fuzzy measure on the set X of criteria is a set function

$$\mu : P(X) \rightarrow [0,1]$$

satisfying the following axioms:

1) $\mu(\Phi) = 0, \mu(X) = 1$
2) $A \subset B \subset X$ implies $\mu(A) \leq \mu(B)$

where $X = \{x_1, \ldots, x_n\}$ is the set of criteria, $P(X)$ is the power set of X, i.e. the set of all subsets of X. Here $\mu(A)$ represents the weight of importance of the set of criteria A. Φ denotes the empty set.

The 2-additive fuzzy measure is a special kind of fuzzy measure. It is defined by

$$\mu(K) = \sum_{i=1}^{n} a_i x_i + \sum_{\{i,j\} \subseteq X} a_{ij} x_i x_j \text{ ,} \qquad (1)$$

where $K \subseteq X$. For any $K \subseteq X$ and $|K| \geq 2$ with x_i=1 if $i \in K$, otherwise x_i=0.

As $\mu_i = a_i$ for all i, the general form for the 2-additive fuzzy measure can be formulated as

$$\mu(K) = \sum_{\{i,j\} \subseteq K} \mu_{ij} - (|K| - 2) \sum_{i \in K} \mu_i \qquad (2)$$

for any $K \subseteq X$ and $|K| \geq 2$. It can be seen from the expression that the 2-additive fuzzy measure is determined by the coefficients μ_i and μ_{ij}.

The 2-additive fuzzy measures need $n(n+1)/2$ coefficients to be defined. The general number of coefficients required to be defined for k-order additive fuzzy measures is $\sum_{j=1}^{k} \binom{n}{j}$.

Fuzzy measures can describe any of the three interactions between two criteria A and B:

3 Synergetic interaction, which can be represented by

$$\mu(A \cup B) > \mu(A) + \mu(B) \text{,} \qquad (3)$$

3 Inhibitory interaction, which can be represented by

$$\mu(A \cup B) < \mu(A) + \mu(B), \tag{4}$$

3 Non-interaction, which can be represented by

$$\mu(A \cup B) = \mu(A) + \mu(B). \tag{5}$$

Classical probability theory can only be applied to the third situation when there is no interaction between two criteria [15].

The Choquet fuzzy integral of a function $f : X \to [0,1]$ with respect to μ is defined by

$$C_\mu(f(x_1),\ldots,f(x_n)) := \sum_{i=1}^{n} (f(x_i) - f(x_{i-1}))\mu(A_i), \tag{6}$$

where $X = \{x_1,\ldots,x_n\}$ is the set of criteria; $P(X)$ is the power set of X, i.e. the set of all subsets of X; $\mu(A)$ is the fuzzy measure representing the importance of the set of criteria A, and $A_i = \{x_i, x_{i+1},\ldots,x_n\}$

The importance index (Shapley value) [16] v_i of element i with respect to a fuzzy measure μ is defined by

$$v_i = \sum_{k=0}^{n-1} \gamma_k \sum_{K \subset X \setminus i, |K|=k} [\mu(K \cup \{i\}) - \mu(K)], \tag{7}$$

where $\gamma_k = (n-k-1)!k!/n!$, X is a set of n elements, $K \subset X$.

Shapley value describes the global importance of each element. It has the property

$$\sum_{i=1}^{n} v_i = 1. \tag{8}$$

Scaled by n, a Shapley value greater than 1 indicates that the element is more important than the average. The Shapley values can also be converted into a percentage to represent the importance of each element

The importance index demonstrates the contributions of individual features to the recognition of a specific class. Based on the importance index, we can determine feature i is more important than feature j if importance index v_i of feature i is great than importance index v_j of feature j.

The interaction index [17] I_{ji} between two elements i and j with respect to a fuzzy measure μ is defined by

$$I_{ij} = \sum_{k=0}^{n-2} \zeta_k \sum_{K \subset X \setminus \{i,j\}, |K|=k} [\mu(K \cup \{i,j\}) - \mu(K \cup \{i\}) - \mu(K \cup \{j\}) + \mu(K)], \tag{9}$$

where $\xi_k = (n-k-2)!k!/(n-1)!$.

The interaction index describes the degree of redundancy of two features to the recognition of a specific class. Feature i and feature j are complementary for recognition of a class from others if the interaction index I_{ji} between two elements i and j is positive. Both the two features should be employed as their combination will enhance the recognition capability. Feature i and feature j are redundant for the recognition of a class from others if the interaction index I_{ji} between two elements i and j is negative. Choosing of either feature i or feature j will have the same effect on final recognition result.

Feature i and feature j are independent for the recognition of a class from others if the interaction index I_{ji} between two elements i and j is zero. In this case, both features have their contributions.

Based on the concept of importance index and interaction index, a feature selection approach for machinery fault diagnosis is proposed. The importance index is used to choose the most important feature set from a number of features. The interaction index is then applied to remove the redundant feature form the set. This approach is expected to optimise the feature group selection for machinery fault diagnosis in different aspects.

3 IDENTIFYING FUZZY MEASURES USING TRAINING DATA SET

As discussed before, the bases of this feature selection method depends on calculation of the importance index and the interaction index with respect to the fuzzy measure μ. This research uses the 2-additive fuzzy measure as it is relatively simple and easy to implement. However, identification of the fuzzy measures is not easy. Here we present a data-driven algorithm for identifying the fuzzy measures.

Identifying fuzzy measures using training data set is an inverse problem of defining fuzzy integrals. Normally the training data set are obtained according to a field expert's knowledge. Here the overall assessment was regarded as Choquet fuzzy integral. The local assessments *f* are based on individual features, see Table 1. In this case, according to the definition of the Choquet integral, identifying the fuzzy integrals means to solve the equation

$$\sum_{i=1}^{n}(f(x_i)-f(x_{i-1}))\mu(A_i)=a_l \quad (l=1,\ldots,m) \tag{10}$$

Table 1

Training data for identifying fuzzy measures

Data point	Local assessments				Overall assessment
1	$f_1(x_1)$	$f_1(x_2)$	...	$f_1(x_n)$	$C_{\mu 1}$
2	$f_2(x_1)$	$f_2(x_2)$	...	$f_2(x_n)$	$C_{\mu 2}$
⋮	⋮	⋮	⋮	⋮	⋮
m	$f_m(x_1)$	$f_m(x_2)$	...	$f_m(x_n)$	$C_{\mu m}$

A heuristic least mean square algorithm (HLMS) is employed to provide an approximate optimal solution of the equation by minimizing the error between the actual Choquet fuzzy integral and the expected Choquet fuzzy integral output

$$E=[C_\mu(f)-y]^2. \tag{11}$$

Conditions such as monotonicity and normalization should also be satisfied when implementing the algorithm. The monotonicity condition can be formulated as

$$\sum_{j\in K}\mu_{ij}-\sum_{j\in K}\mu_j-(n-2)\mu_i\geq 0 \tag{12}$$

for $\forall i\in X$, $K\subseteq X\setminus i$, where $|X|=n$.

The normalization condition can be formulated as

$$\sum_{\{i,j\}\subseteq X}\mu_{ij}-(n-2)\sum_{i\in X}\mu_i=1 \tag{13}$$

4 FEATURE GROUP OPTIMISATION BASED ON FUZZY MEASURES

The proposed optimisation method for feature group optimization can be implemented in four steps (Figure 1). First we need to obtain the fault knowledge, which is represented by a training data set. Based on the training data set, a specific kind of fuzzy measure is extracted. The interaction index with respect to the fuzzy measures, which is used to remove redundant features from a group of feature set, is then computed. The last step of the approach is to compute the importance index to choose the most important features from a number of features. Sometimes we can change the order of the last two steps.

To obtain the fault knowledge, fuzzy *c*-means (FCM) clustering analysis is employed. It is used by inputting both individual and combinations of fault features to acquire both the local and overall assessment of an object. The overall assessment is used as Choquet fuzzy integrals and the local assessment is used as function f to describe the contributions of individual features to the object.

The core part of the heuristic least mean square algorithm (HLMS) for the 2-additive fuzzy measure identification adopts the gradient descent theory

$$\mu_{new} = \mu_{old} - \alpha \frac{E}{E_{\max}} (x_{n-i} - x_{n-i-1}) \qquad (14)$$

where i denotes the ith element of feature vector in ascending order; $\alpha \in [0,1]$ is a constant, which is known as learning rate; E is the error between actual and expected output.

Once the fuzzy measures are determined, the interaction index of every pair of features and importance index of each feature can be computed. Based on the importance index and interaction index, the less important and redundant features will be eliminated.

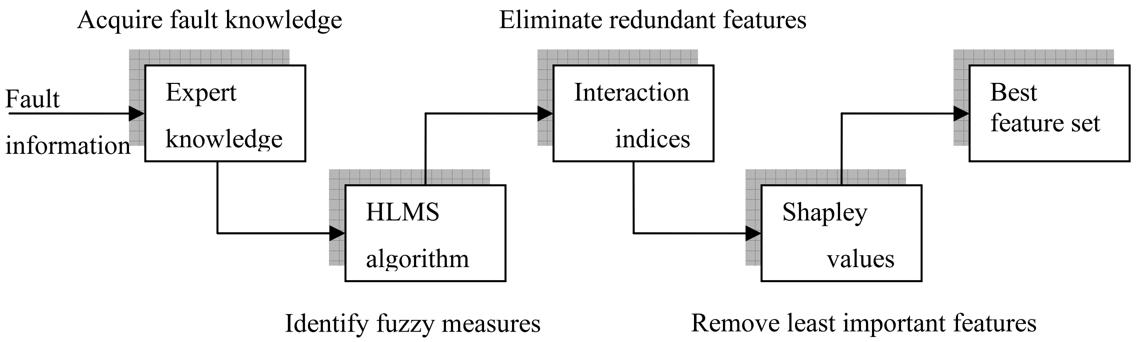


Figure 1 Feature group optimization based on fuzzy measures

5 EXPERIMENT RESULTS

An experiment on rolling element bearing faults was conducted using a machinery fault simulator made by SpectraQuest Inc to evaluate the proposed fault diagnostic method. The data acquisition system included an amplifier PCB 482A20, a filter KROHN-HITE 3202, a NI A/D card Daq6062E with a connector BNC2120, two vibration sensors IMI 608A11 and a Toshiba laptop computer. The software used was LabView. The cutoff frequency was 5k Hz and the sampling rate 20k Hz. Simulated defects included inner race fault, outer race fault, ball fault, looseness of the bearing housing and different fault combinations under different loads and rotating speeds.

Four time-domain features were extracted from the vibration signals, including Root Mean Square (RMS), maximum value, kurtosis and crest factor. The feature values were normalized into range [0, 1] to meet the requirement of the employed fuzzy method input.

For machinery fault diagnosis, although quite a few features are available, their degrees of contribution to the diagnostic process are usually different. The purpose of this research is to keep the minimum number of features while still maintain the diagnostic accuracy. In this specific scenario, we managed to choose two most suitable features from the group of four to validate our proposed method. Here we used x_1, x_2, x_3, and x_4 to denote RMS, maximum value, kurtosis and crest factor, respectively.

The fault knowledge is normally obtained empirically or according to expert field knowledge. This research used fuzzy *c*-means clustering algorithm as an expert system to acquire the fault knowledge from the original signal. The fault knowledge included local and overall assessment of a specific fault based on different features.

Once the fault knowledge extracted, the 2-additive fuzzy measure was identified. Table 2 shows the identified fuzzy measures for bearing outer race fault. Table 3 shows the Shapley values of each feature for bearing outer race fault.

Table 2

The 2-additive fuzzy measures on the power set of features

Subset	$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_4\}$		
Fuzzy measure	0.2975	0.2623	0.3350	0.3268		
Subset	$\{x_1,x_2\}$	$\{x_1,x_3\}$	$\{x_1,x_4\}$	$\{x_2,x_3\}$	$\{x_2,x_4\}$	$\{x_3,x_4\}$
Fuzzy measure	0.5680	0.6375	0.5000	0.5784	0.5310	0.5987
Subset	$\{x_1,x_2,x_3\}$	$\{x_1,x_2,x_4\}$	$\{x_1,x_3,x_4\}$	$\{x_2,x_3,x_4\}$		
Fuzzy measure	0.8890	0.7124	0.7769	0.7840		

It can be seen from Table 3 that all the interaction index involving feature x_4 are negative. This indicates that Crest factor is a redundant feature. It can also be seen from Table 4 that Crest factor plays the least important role in diagnosing bearing fault in this case. As such, this feature should be removed from the feature group. Table 4 shows that the second least important feature is the Maximum value. Although it is complementary with other features, it was eliminated from the set, as a feature subset of only two elements was pursued in this case.

Table 3

Interaction index

Feature set	$\{x_1,x_2\}$	$\{x_1,x_3\}$	$\{x_1,x_4\}$	$\{x_2,x_3\}$	$\{x_2,x_4\}$	$\{x_3,x_4\}$
Interaction index	0.0720	0.0706	-0.0425	0.0187	-0.0837	-0.0532

Table 4

Importance index of different feature

Feature	x_1 (RMS)	x_2 (Maximum)	x_3(Kurtosis)	x_4(Crest factor)
Importance index	0.2493	0.2353	0.3039	0.2115

As a result, the two features of Crest factor and Maximum value were eliminated from the original feature set {*RMS, maximum value, kurtosis, crest factor*} based on important index and interaction index with respect to the 2-additive fuzzy measures. The best feature subset for diagnosing the bearing fault, {*RMS, kurtosis*} was eventually obtained.

6 DISCUSSIONS

Fuzzy measure and fuzzy integral techniques are able to represent importance of criteria and certain interactions among them. Based on this concept, a feature group optimization approach for machinery fault diagnosis is proposed. For rolling element bearing fault diagnosis, the feature subset {*RMS, kurtosis*} was selected from a feature set of four {*RMS, maximum value, kurtosis, crest factor*}. To validate the result, fuzzy *c*-means clustering was used for comparison. Table 5 illustrates the average values of the membership degree of all features of the bearing fault. By comparing Table 4 with Table 5, it can be seen that the two results are in full agreement with each other.

Table 3

Average value of membership degree of features

Feature	RMS	Maximum	Kurtosis	Crest factor
Average value	0.9523	0.9118	0.9524	0.1729

Even from the 2-additive fuzzy measures, some initial signs can be observed. The fuzzy measure for $\{x_1, x_3\}$ is greater than other fuzzy measures of the two-element subset. The fuzzy measure of a three element set which contains x_1 and x_3 is also greater than other fuzzy measures of the three-element subset. The initial assumption was further proven by using the importance index and the interaction index.

7 CONCLUSION

This paper presents a novel fuzzy measure based feature optimization approach for machinery fault diagnosis. This method can produce an optimal feature set in terms of the importance index and the interaction index. In the case study, a two feature subset {*RMS, kurtosis*} was selected from a feature set of four {*RMS, maximum value, kurtosis, crest factor*} for rolling element bearing fault diagnosis. Experimental results show that the proposed feature selection approach has great potential in efficient and effective fault diagnosis.

8 REFERENCES

[1] Xu, Z., et al., Diagnosis Principle and Application for the Fault Information. 2000: China Defence Industry Press.

[2] Sun, J., Pattern Recognition. 2002, Changsha: Defence Technology University.

[3] Sommol, P. and P. Pudil, Feature selection toolbox. Pattern Recognition, 2002. 35: p. 2749-2759.

[4] Pudil, P., J. Novovicova, and P. Somol, Feature selection toolbox software package. Pattern Recognition Letters, 2002. 23: p. 487-492.

[5] Schulerud, H. and F. Albregtsen, Many are called, but few are chosen. Feature selection and error estimation in high dimensional spaces. Computer Methods and Programs in Biomedicine, 2004. 73: p. 91-99.

[6] Chen, X.-W., An improved branch and bound algorithm for feature selection. Pattern Recognition Letters, 2003. 24: p. 1925-1933.

[7] Yu, B. and B. Yuan, A more efficient branch and bound algorithm for feature selection. Pattern Recognition, 1993. 26: p. 883-889.

[8] Gader, P.D., M.A. Mohamed, and J.M. Keller, Fusion of handwritten word classifiers. Pattern Recognition Letters, 1996. 17: p. 577-584.

[9] Michel, The representation of importance and interaction of features by fuzzy measures. Pattern Recognition Letters, 1996. 17: p. 567-575.

[10] Kuncheva, L.I., J.C. Bezdek, and R.P.W. Duin, Decision templates for multiple classifier fusion: an experimental comparison. Pattern Recognition, 2001. 34: p. 299-314.

[11] Mirhosseini, A.R. and H. Yan, Human face image recognition: an evidence aggregation approach. Computer Vision and Image Understanding, 1998. 71: p. 213-230.

[12] Pham, T.D. and H.Yan, Color image segmentation using fuzzy integral and mountain clustering. Fuzzy sets and systems, 1999. 107: p. 121-130.

[13] Kwak, K.-C. and W. Pedrycz, Face recognition using fuzzy integral and wavelet decomposition method. IEEE transactions on systems, man, and cybernetics--Part B: cybernetics, 2004. 34: p. 1666-1675.

[14] Auephanwiriyakul, S., J.M. Keller, and P.D. Gader, Generalized Choquet fuzzy integral fusion. Information Fusion, 2002. 3: p. 69-85.

[15] Klir, G.J., Z. Wang, and D. Harmanec, Constructing fuzzy measures in expert systems. Fuzzy sets and Systems, 1997. 92: p. 251-264.

[16] Shapley, L.S., A value for n-person game. Contributions to the theory of games, ed. H.W. Kuhn and A.W. Tucker. Vol. 2. 1953, Princeton, New Jersey: Princeton University Press. 307-317.

[17] Murofusion, T. and S. Soneda. Techniques for reading fuzzy measures (3): interaction index. in Proceedings of the 9th fuzzy systems symposium. May 1993. Sapporo, Japan.

TECHNIQUES IN DEVELOPING ECONOMIC DECISION MODEL COMBINING ABOVE RAIL AND BELOW RAIL ASSETS

V. Reddy[1], G. Chattopadhyay[2], D. Hargreaves[3], P. O. Larsson- Kråik [4]

[1, 2 & 3] *Queensland University of Technology School of Engineering Systems,*

2, George Street P.O Box 2434 Brisbane QLD 4001, AUSTRALIA

v.reddy@qut.edu.au

g.chattopadhyay@qut.edu.au

d.hargreaves@qut.edu.au

[4] ***Swedish National Rail Administration, Technical Department,***

P.O. Box 43 SE-971 02 LULEÅ, SWEDEN.

per-olof.larsson-kraik@banverket.se

ABSTRACT:Many of the rail infrastructure owners take ad-hoc decisions in inspection, lubrication and grinding strategies for controlling rolling contact fatigue (RCF) and wear. Limited research has been conducted in this area for economic effectiveness combining above rail and below rail assets. There is need to focus on damage of rail caused due to wheel and damage of wheel caused due to rail and total cost of rail and wheel for optimal replacements, maintenance and penalty costs to rail or train owners. This paper focuses on techniques in developing integrated economic decision model combining above rail and below rail assets for operational and maintenance strategies. These techniques able to analyse the effectiveness of operational and maintenance related decisions combining rail and wheel considering various scenarios.

KEYWORDS: Wear-Fatigue-Lubrication-interaction, Grinding, Inspection and Rail-wheel replacements

1 1. INTRODUCTION

Many of the rail infrastructure owners take ad-hoc decisions in inspection, lubrication and grinding strategies for controlling rolling contact fatigue (RCF) and wear. Rail and wheels deteriorate due to rail-wheel interaction in the form of wear, plastic deformation and fatigue. This includes rolling contact fatigue cracks, loss of profile and development of corrugations. Rail grinding is widely used to control some of these problems. The appropriate rail-grinding interval depends on the rail metallurgy, track curvature, axle loads. Magel and Kalousek (2002) recommended an interval of 8-12 MGT (280-300 BHN standard carbon rail) and 12-25 MGT (360-380 BHN premium rail) for sharp curves. Decisions on rail grinding to reduce rolling contact fatigue (RCF), lubrication to reduce wear and rail replacements to reduce risk of derailments are taken on the basis of rail life data. Similar decisions are taken for rolling stock based on rate of degradation and actual condition at the time of inspection. Ishida et al., (2003) studied rail grinding in practice to determine grinding interval (how frequently the grinding should be conducted) and grinding depth (how much surface material should be removed). Rail players around the world take decisions based on Visual Qualitative Checks (VQC), Non Destructive Testing (NDT), assumed Million Gross Tonnes (MGT) for axle loads and Traffic Density (TD). This is a slow, time consuming and costly process of decision making leading to a risky operating condition in between inspections.

Rail-wheel lubrication reduces wear rates for rail infrastructure owners and increases fuel savings for rail operators. Three lubrication systems used generally are hi-rail, wayside and on-board lubrication systems. Hi-rail lubrication system uses a specially designed mobile truck for the grease application from the nozzle as a thin bead along the rail gauge face. Wayside lubrication system grease is applied at track when the lubricator is activated either mechanically or electronically by passing wheels. On board lubrication system, the lubricator is mounted on the locomotive and the lubricant is applied using a spray system to the locomotive wheel flanges. By using one or a combination of the above systems, railroads are able to achieve significant savings in fuel and cost of wheel/track maintenance. However, there are some adverse effects of using excessive lubricant. These are wastage, loss of locomotive traction due to presence of lubricant on the top of rail and environmental concerns due to underground water contamination (Pandey et. al., 2000).

MTR Corporation in conjunction with BHP Research developed an economic/technical Rail Management Model to optimise the rail maintenance and replacement strategies for the MTR system under specific operating conditions (Keefe and Soleiman, 1994). The improved rail-wheel contact conditions significantly reduced rail replacements due to gauge corner shelling. It is found that rail replaced during the year 2000 was approximately 85% less than the rail replaced in the year 1992 whilst the annual gross tonnage has increased approximately by 10% (Keefe and Ravitharan, 2001). Achieving optimised rail and wheel profiles influence the contact mechanics and significantly affect wheelset and steering dynamics. It has helped rail infrastructure owners in reducing cracking and shelling defects and excessive wear and corrugation, and improving vehicle performance. A worn wheel profile is the result of all the rail profiles it encounters on a particular route and a worn rail is the result of all of the wheel profiles that pass over it.

This paper focuses on techniques in developing integrated economic decision model combining above rail and below rail assets for operational and maintenance strategies. Outline of this paper as follows: Section 1 provides introduction with an overview of existing rail-wheel inspection and maintenance. Section 2 focuses on comparative evaluation of these techniques. Section 3 proposes economic model combining above rail and below rail assets for operational and maintenance decisions. The concluding section discusses summary and scope for future work.

2 COMPARATIVE EVALUATION OF EXISTING TECHNIQUES

2.1 Rail and wheel wear

Wear is loss of material from the surface of rail and wheel due to rolling and sliding contact. MiniProf Rail profile system is used to measure the profiles just before and after preventive rail grinding as shown in Figure 1. Further information on the MINIPROF system can be found in Esveld and Gronskov (1996). The accuracy of the MINIPROF system is in the order of ± 0.015mm.

Figure 1: Rail Profile Measurement using MINIPROF Rail profile system [MINIPROF, Greenwood, Denmark]

Kalousek and Magel (1997) introduced magic wear rate. The boundary between sufficient and insufficient wear – the optimal, or "magic", wear rate – represents a state of balance in which the development of fatigue is arrested by wear or metal removal. The magic wear rate is achieved when the surface material wears just enough to prevent surface fatigue cracks from propagating and causing fatigue failures. Rail owners are still looking for better wear measurement to determine wear rate for rail and wheel to reduce cost of replacements and risk of derailments. Epp and Mutton (2002) explains that optimisation of wheel-rail interface is a continuous process. Ongoing review and refinement of wheel and rail profiles are required as long as there are changes in vehicle types and traffic conditions.

2.2 Rolling Contact Fatigue (RCF)

There are different approaches to analyse fatigue crack initiation such as, the defect-tolerant approach and the total-life approach. The total-life approach estimates the resistance to fatigue crack initiation based on nominally defect-free materials/ components (Ringsberg, 2001). The strain-life approach together with the elastic-plastic FE model is used to predict:

- The position for greatest fatigue damage
- The orientation of crack planes
- The fatigue life to crack initiation due to both low-cycle fatigue and ratchetting.

Yokoyama et al., (2002) investigated rolling contact behaviour in pearlitic and bainitic steel rails with various angle of attack in the twin-disc type RCF test machines.

2.3 Rail-Wheel Lubrication Strategies

Railway flange lubrication is used to reduce wear of the rail and wheel which is particularly severe at curves. The American Association of Railroads (AAR) estimated that the wear and friction occurring at the wheel/rail interface of trains due to ineffective lubrication costs American Railways in excess of US $ 2 billion each year. Currently, the largest expenditure faced by the railroad industry is the rail maintenance and replacement. Application of lubricants at the wheel rail interface dramatically reduces the rail track degradation and fuel consumption (Sid and Wolf, 2002). Rail life has been increased by factor of two and wheel life by factor of five (Queensland Rail, Australia) using appropriate lubrication. Spoornet (South Africa) has reported that rail life was increased from 27 MGT to up to 350 MGT depending on curve radius. Canadian Pacific (CP) rail indicated that rail life improvement by 110% using effective lubrication. Experiments on the Olympic park loop, NSW (Australia) on a 200 m radius curve indicated that flange wear rate is reduced from 0.36 mm/day (life of 0.2 year) to 0.006 mm/day (life 3.5 years). HKMTRC (Hong Kong) reported a cost saving of £ 783,000 per year on wheel and rail maintenance on the solid lubricant lines. Eurostar conservatively estimates that this saves £ 1,000,000 per year on maintenance and wheel replacement costs with effective lubrications (Larke, 2003). ERL/Malaysia has recorded data on lubrication shown in Table 1.

Table 1

Cost of Lubrication strategies (Larke, 2003)

Track/Vehicle condition	Wheel life in km	Wheel life in weeks	*Annual wheel cost in (£)*
No lubrication	170,000	20	*1.6 million*
Rail lubrication	300,000	35	*825,000*
Vehicle lubrication	1,000,000	118	*250,000*
Target	*1,500,000*	*177*	*170,000*

Typical lubricant costs in the UK are between £ 2.50/kg to £ 6/kg. In the filed measurement Norfolk Southern, the electric lubricators achieved 107% improvement in lubricant dispersion along the track, a 67% reduction in lubricant usage and a 57% of reduction in wastage (Larke, 2003). Typical grease usage data, lubricant costs and maintenance costs are shown in the Table 2.

Table 2

Lubrication costs to rail players (Larke, 2003)

Railway	Quantity(tonnes/yr)	Lubricator(£/yr)	Lubricant(£/yr)	Costs (£/yr)
Spoornet	200	125,000	134,000	259,000
IBCV	Not known	325,000	279,000	604,000
HKMTR	0.3 (depots)	5,600	550	6,150
Eurostar	1.1	Not known	Not known	70,000
Banverket	20	Not known	31000 - 62000	Not known
DSB	25	95,000	100,000	195,000
SNCB	40	1,000,000	479,000	1,479,000

Magel and Kalousek (2002) discussed the optimal rail wheel profiles. Ishida et al., (2003) discussed about wheel rail interface problems such as rolling contact fatigue (RCF), squat defects, grinding, corrugations, and lubrication of Japanese Railways (JR). Preventive grinding is currently most reliable and effective method of reducing the occurrence of RCF defects. In Japan grinding started in the 1970s and only about 100 km of affected track was ground every year. In 1995 two rail grinding cars were introduced. Since that time rail grinding has been carried out more than 1000 km of track per year (Tada, 1999). It is found that the number of squats has been steadily decreasing as a result of grinding. The target has been a grinding thickness of 0.08 mm/pass and a grinding interval of 40 MGT, which is similar to 0.1 mm/50 MGT. In this study it is found that the effect of preventive grinding has not been scientifically evaluated. It is essential to obtain optimal grinding period and grinding amount (the thickness of material removed from the rail surface) to control such defects.

3 ECONOMIC MODEL COMBINING ABOVE RAIL AND BELOW RAIL

The rail and wheel interface produces wear and fatigue. Unbalanced or damaged wheels have adverse effect on rails. Similarly, uneven wear rate of rails or worn out rails have adverse effect on wheels. Achieving balance and charging damage caused by rails to rail infrastructure operators or caused be wheels to train operators is a big challenge. It is important to predict wear and fatigue in both rails and wheels to maintain the balance. Wheel-rail interaction model proposed by Mutton (2001) found that 12 m of rail (head hardened) track length is required to wear out one wheel for a worn wheel but approximately 17.5 m is required to wear out one wheel for a new (head hardened) rail. Similarly 20 m standard carbon rail is required to wear out one worn wheel and approximately 25 m required to wear out one new wheel. The values changes with the curve radius increase of axle loads and train speed and traffic density.

A conceptual rail wheel damage interaction could be any of the following ones:

$$y_i(w) = a_i e^{-b_i r} \tag{1}$$

Where w is wheel wear loss, r rail wear loss,

a and b are parameters based on rail-wheel materials and conditions.

$$y_i(w) = a_i + b_i r^1 + c_i r^2 + d_i r^3 \tag{2}$$

Where a, b, c, and d are parameters based on rail-wheel materials and conditions.

r is rail wear loss for different rail materials (such as head hardened and standard carbon), rail size, and condition (new and worn rail and wheels)

$$y_i(w) = a_i(r)^{b_i} \qquad (3)$$

Where w is wheel wear loss, r rail wear loss,

a and b are parameters based on rail-wheel materials and conditions

It is also possible to model rail loss for known wheel material and condition for certain operating conditions.

Table 3

Rail-wheel damage interaction data

Serial No	Scenario	Rail wear loss (mm^2)	Rail track length (m)
1	Head hardened (HH) rail and worn-n-wheels	10	12
		20	6
		30	4
		40	3
		50	2.5
2	Head hardened (HH) rail and new wheels	10	17
		20	8.5
		30	5.5
		40	4.25
		50	3.25
3	Standard Carbon (SC) rail and worn-n-wheels	10	20.5
		20	10.5
		30	7
		40	5.25
		50	4
4	Standard Carbon (SC) rail and new wheels	10	26
		20	14.10
		30	9.75
		40	7
		50	5.75

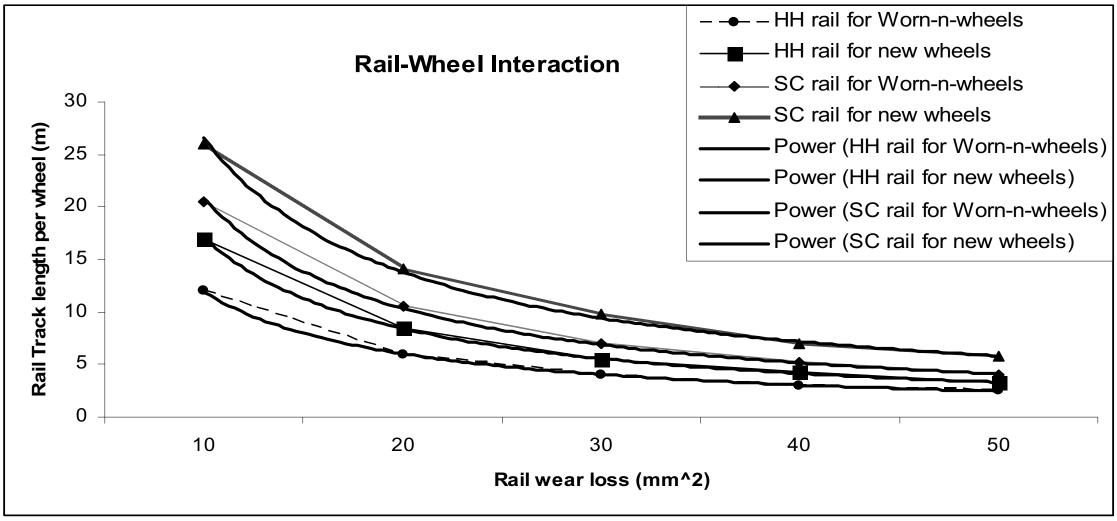


Figure 2: Above rail and below rail damage interaction.

Figure 2 shows the analysis of above rail and below rail model for rail wheel interaction. It is found from the analysis that Figure 2 better captures the real life case. It is also found that standard carbon (SC) rails have less adverse effect on wheel life damage compared to head hardened rails. Figure 3 shows the event tree analysis of rail-wheel degradation.

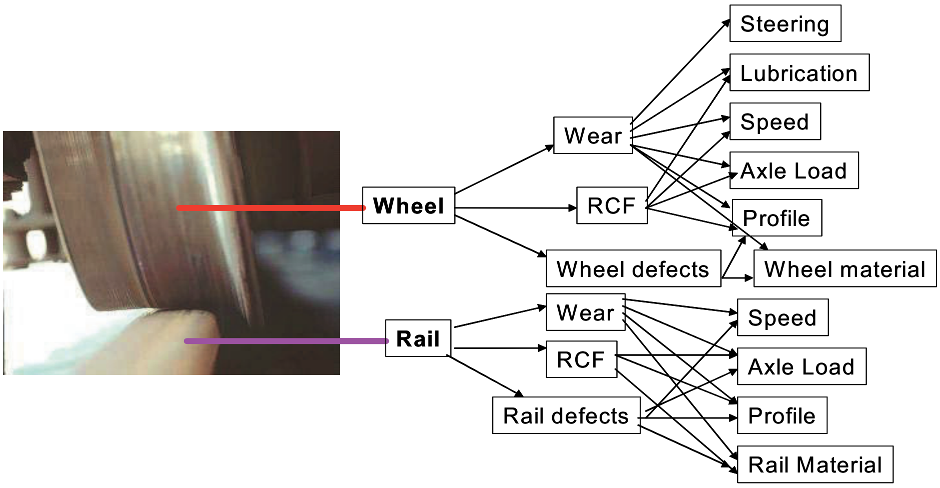


Figure 3: Factors influencing rail-wheel degradation (Chattopadhyay et al., 2004)

There is a need for development of rail-wheel interaction models for reducing risk of derailment and to improve safety and reliability of rail life. Figure 4 shows proposed rail-wheel damage interaction model. This model can be used to estimate the costs due to rail wear loss and wheel wear loss as a result of rail-wheel damage interaction.

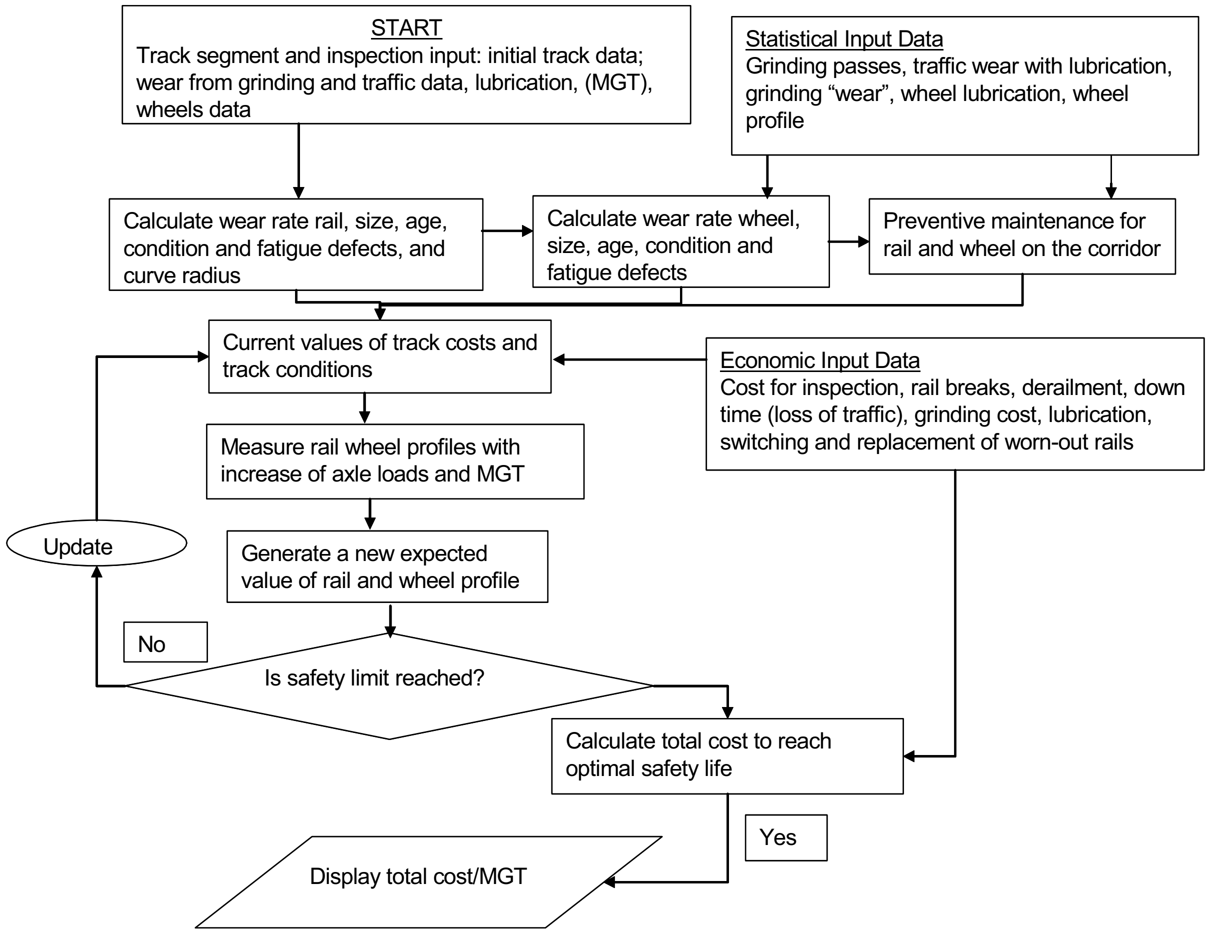


Figure 4: Proposed model for rail-wheel damage interaction model

4 SUMMARIES AND SCOPE FOR FUTURE WORK

This paper focused on techniques in developing integrated economic decision model combining above rail and below rail assets. Rail-wheel damage interaction models proposed based on rail wear loss and wheel wear loss considering new and worn wheels for various rail materials and wheel conditions. Proposed models were developed based on Queensland Rail (QR) internal report. Further these models could be illustrated and validated with laboratory and real life data. There is scope for future work in development of rail-wheel damage interaction model under various operating conditions and traffic types. This model can be used as a framework for penalty costs related to rails damaging wheels or wheels damaging rails and who pays for what. Authors are currently working in this area and results would be published in the future.

5 REFERENCES

[1] Cannon, D.F., Edel, K.O., Grassie, S. L. & Sawley, K. (2003) Rail Defects: an overview. *Fatigue & Fracture of Engineering Materials & Structures*, 26(10), 865-887.

[2] Chattopadhyay, G., Reddy. V., Hargreaves, D. and Larsson, P. O., (2004) Integrated Rail & Wheel Maintenance Model for Cost sharing by Rail Players. *Proceedings of the 3rd International Conference on Vibration Engineering and Technology of Machinery and the 4th Asia-Pacific Conference on System Integrity and Maintenance (VETOMAC-3 & ACSIM-2004)*, 6th – 9th December 2004, New Delhi, India, Vol I, 170-177, 81-7764-700-8.

[3] Clark R., (2004) Rail flaw detection: overview and needs for future developments. *NDT&E International* 37, 111–118.

[4] Epp, K. J., Mutton, P. J., (2002) Wheel/Rail Interaction: Current “State of The art” in the Australian Railway Industry. *Conference on Railway Engineering*, Wollongong, 10-13 November 2002.

[5] Esveld, C. and Gronskov, L., (1996) MINIPROF wheel and rail profile measurement. *2nd Mini-conference on contact mechanics and wear of rail/wheel systems*, Budapest, July 29–31, 1996.

[6] Fletcher, D. I. & Beynon, J. H., (2000) The effect of contact load reduction on the fatigue life of pearlitic rail steel in lubricated rolling-sliding contact. *Fatigue & Fracture of Engineering Materials & Structures* 23(8), 639-650.

[7] Ishida, M., Akama, M., Kashiwaya, K. and Kapoor, A., (2003) The current status theory and practice on rail integrity in Japanese railways – rolling contact fatigue and corrugations. *Fatigue & Fracture of Engineering Materials & Structures*, 26(10), 909-919.

[8] Keefe, R. & Soeleiman, S., (1994) The Development and Application of a Technical/Economical Rail Management Model to Hong Kong’s Mass Transit Railway. Tenth International Rail Track Conference, Brisbane, September, pp. 131-143.

[9] Keefe, R. & Ravitharan, S.S.S., (2000) Managing Rail Deterioration, Railway Engineering 2001, London, April.

[10] Kalousek, J., Magel, E., (1997) Achieving a balance: The “magic” wear rate. Railway Track & Structure, pp. 50-52.

[11] Larke R, (2003) Survey Of Wheel / Rail Lubrication Practices. Research and Development Programme, Rail Safety Standard Board, UK, Report no: ITLR-T12442-001.

[12] Magel, E., Kalousek, J., (2002) The application of contact mechanics to rail profile design and rail grinding. Wear, 253(1-2), 308-316.

[13] Mutton, P. J., (2001) Rolling Contact Fatigue In Railway Wheels Under High Axle Loads. BHP Institute of Railway Technology, BHP Research Report.

[14] Pandey, L. M., Saxena, D., Sayanna, E., Mookken, R. T., Srivastava, S. P., and Bhatnagar, A. K., (2000) Railway Wheel Flange Lubrication A New Approach. Indian Oil Corporation Limited, R & D Centre, Sector 13, Faridabad.

[15] Ringsberg, J.W., (2001) Life prediction of rolling contact fatigue crack initiation. International Journal of Fatigue, 23, 575-586.

[16] Sid D, Wolf, E (2002) Transportation FOR THE 21ST CENTURY. Trac Glide Top-of-Rail Lubrication System, Report from Department of Energy, USA.

[17] Tada, Y. (1999) Rail grinding method in Shinkansen. Sinsenro, 53(8) 4-7.

[18] Yokoyama, H., Mitao, S., Yamamoto, S., and Fujikake, M., (2002) Effect of the angle of attack on flaking behaviour in pearlitic and bainitic steel rails. Wear, 253(1-2), 60-66.

DEVELOPMENT OF WEAR-FATIGUE-LUBRICATION-INTERACTION MODEL FOR COST EFFECTIVE RAIL MAINTENANCE DECISIONS

V. Reddy[1], G. Chattopadhyay[2], D. Hargreaves[3], P. O. Larsson- Kråik [4]

[1, 2 & 3] *Queensland University of Technology, School of Engineering Systems,*

2, George Street P.O Box 2434 Brisbane QLD 4001, AUSTRALIA

v.reddy@qut.edu.au

g.chattopadhyay@qut.edu.au

d.hargreaves@qut.edu.au

[4] *Swedish National Rail Administration, Technical Department P.O. Box 43 SE-971 02 LULEÅ, SWEDEN*

per-olof.larsson-kraik@banverket.se

Abstract: Prediction of operational risk and enhancement of reliability and safety for rail network are important to rail infrastructure providers and train operators. Rail defects leading to rail breaks and derailments cost huge amount due to cancelled / delayed traffic, emergency maintenance, loss of assets, loss of revenues and liability compensations due to down times. Wear occurring at the wheel/rail interface as a result of ineffective lubrication costs huge amount to rail infrastructure providers per year. Cost of rail-wheel maintenance and replacements are reduced significantly with effective lubrication. Studies on interaction of factors behind wear, fatigue, lubrication and grinding problems are required to monitor those factors and finding out cost effective technological solutions in predicting and reducing operational risks. This paper focuses on development of wear-fatigue-lubrication-interaction model for cost effective rail maintenance decisions.

Key Words: Wear-Fatigue-Lubrication-interaction, Grinding, Inspection and Operational risks

1 INTRODUCTION

Wear and fatigue of rail and wheel are growing concerns to rail infrastructure players. These are main concerns of heavy haul operators and on heavy traffic density lines. Material loss in wear occurs in the form of debris, transfer of material from one surface to another by adhesion or by local plastic deformation. In the rail-wheel contact three wear regimes are identified: they are mild, sever and catastrophic wear [1]. Rolling contact fatigue (RCF) occurs due to repetitive load of wheels on track. Some of the adverse effects are squats, shelling and head checks. Problems at curves and switches due to frequent contacts near gauge corners occurs due to reduced contact area and high slip resulting severe contact stresses [2]. Rail grinding is used to remove fatigue cracks on rail surface and to optimise rail profile. Literature shows that increase in rail grinding interval frequency has substantially increased rail maintenance costs. Rail infrastructure owners around the world have been using lubricant routinely to reduce noise, wear and friction. Effective application of lubrication reduces energy consumption and also extended rail life up to two to three years even in heavy traffic areas. Lubrication commonly used by rail infrastructure providers and rail operators are way side lubrication, on board lubrication and hi-rail lubrication [3]. Lubrication has influence on accelerating fatigue cracks on rail surface under increased axle loads and speed in heavy traffic areas. Lubricant entrapment has adverse effect on initiation and propagation of fatigue cracks leading to rail breaks and catastrophic derailments. Wear can be used to reduce the initiation and propagation of rolling contact fatigue (RCF) cracks. The severe or catastrophic (increase in material loss) wear has significant reduction in rail life. Therefore, understanding wear-fatigue-lubrication interaction is important for cost effective maintenance decisions which results in reduced operational risks and costs.

The outline of this paper: Section 1 introduces rail wear-fatigue-lubrication interaction. Section 2 focuses on overview of wear-fatigue-lubrication interaction. In Section 3 Integrated model is proposed for wear-fatigue-lubrication interaction. Section 4 proposes cost model for effective maintenance decisions. The concluding section discusses summary and scope for future work.

2 OVERVIEW WEAR-FATIGUE-LUBRICATION INTERACTION

Rail wear rate depends on increase of million gross tonnes (MGT), axle loads, train speed, direction, and curve radius and traffic density. Wear limit varies with rail operating and environmental conditions. Many rail infrastructure owners have developed their own wear limit standards based on experience with these factors to control rail wear rate and to reduce risk of low rail derailments. Effective application of lubrication reduces slip and creep force regime between flange and gauge corner. It also reduces traction, material wear and flow and fatigue effects [4].

Wear-fatigue-lubrication interaction has currently received attention from rail infrastructure owners. Recent study shows that huge amount of lubrication are wasted annually in the effort to control wear. The American Association of Railroads estimates that the wear occurring at the wheel/rail interface as a result of ineffective lubrication costs in excess of $US 2 billion per year. Eurostar estimated saving of £ 1,000,000 per year on maintenance and wheel replacement with effective lubrication [5]. The largest expenditure is rail maintenance and replacement [6]. Lubrication reduces traction stress on rail gauge face and on the gauge corners. This increases the wheel loading cycles which influences rolling contact fatigue (RCF) defects. Preventive rail grinding reduces fatigue initiated surface cracks. However, RCF in combination with lubrication influences rail life significantly. When lubrication is applied at rail gauge face, the top of the high rail and low rail remain dry. This increases angles of attack compared to that for dry gauge face. It results in increased lateral forces and contact stress on the gauge corner of the high rail leading to gauge corner shelling of the high rail [7].

Cannon and Pradier [8] analysed various surface initiated cracks in railheads. Head checks (HC) occur near the (rail) gauge corners of the curves and crossings due to repeated plastic deformation and consecutive accumulated damage at the surface of railhead due to rolling contact fatigue (RCF). These accumulated head checks can cause gauge corner break up to a depth of several millimetres. In rare cases, the cracks propagate in a transverse direction with a complete fracture of the rail, known as rail break. Existing theories have limitations in accurately predicting rolling contact fatigue (RCF) and initiation of crack leading to rail breaks [9]. Sawley and Kristan [10] conducted small- and full scale tests to potential resistance of rolling contact fatigue damage. It was found that wear performance of bainitic rail steel depends considerably on test conditions, the indication was that bainitic steel rails can have significantly better rolling contact fatigue performance compared to pearlitic rails. Lubrication has adverse effect on the pressure distribution along surface crack and helps to propagate influencing rail break and risk of derailment. Operational risks are further influenced by number of axle passes, curve radius, rail-wheel profile and material, hardness of material, rail/wheel interaction, rail-wheel grinding intervals, and lubrication at rail-wheel interface and maintenance decisions. If not monitored and controlled properly these can lead to rail defects, rail breaks and derailments. These problems can cause huge loss of revenue, disruption of service, resulting damage of assets, and loss of lives. In 2000, the cost of the Hatfield, UK accident was £730 million for repairs and compensation payments. The main cause was RCF [11].

3 INTEGRATED MODEL FOR RAIL WEAR-FATIGUE-LUBRICATION-INTERACTION

Lubrication has a substantial impact on rail life by reducing wear at sharp curves resulting in improvement of reliability and safety of railroads. However, fluid lubricants contribute to the development of rolling contact fatigue (RCF) cracks by mechanisms, e.g. crack face friction modification and fluid entrapment shown in Figure 1. The possibility therefore exists that while application of a fluid lubricant may reduce rail and wheel wear, the lubricant may lead to rail RCF failures. Dry solid lubricants have advantage to avoid this problem [12].

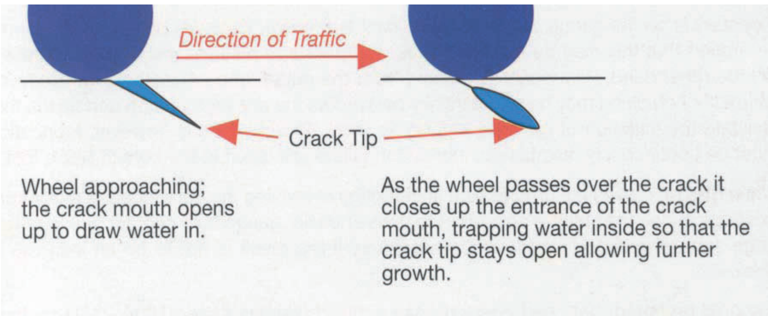


Figure 1: Fluid entrapment in rolling contact fatigue cracks of rail surface [13]

Effective lubrication can be achieved only with

- suitable weather and environmental friendly and improved efficient lubricant
- appropriate position of way side lubricators
- effective maintenance of lubricators
- effective application (distribution) of lubricant along the rail and wheel surface
- balanced (even) distribution of lubricant throughout the curve length of rail
- cut-off point for curve radius that need to be to lubricated above recommended curve radii
- dismantling of blades from the track before grinding and reinstalling after grinding can lead to additional maintenance costs
- exploring the possibilities of hi-rail and on-board lubrication for cost effectiveness application of lubrication
- applying cheap lubricant may not be an effective (optimal) solution, which inturn lead to additional maintenance costs of lubricators.

A lubricated wheel/rail contact operates with boundary or mixed film lubrication. The difference between these regimes is shown in Figure 2.

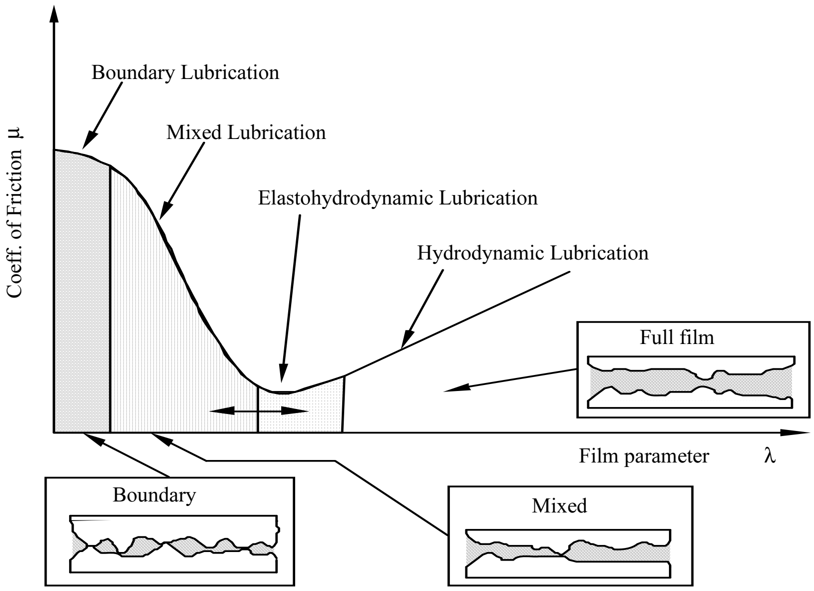


Figure 2: Coefficient of friction varies with the film thickness parameters [14]

Therefore, there is a need to develop a lubrication decision model for benchmarking lubrication for various curve radii considering extreme environmental conditions. Rail grinding has been effective solution to reduce rolling contact fatigue cracks and to correct rail profile. Rail grinding is a costly maintenance activity used to prevent fatigue cracks, rail breaks, failures and catastrophic derailments. Rail players around the world also using balanced (sufficient) wear to control rolling contact fatigue. Figure 3 shows the twin disc test rig developed at QUT and was being used to conduct wear and lubrications tests.

Figure 3: Twin disc test rig developed at QUT

Besuner et al., [15] pointed out limitations of rail life prediction models using MGT because it does not differentiate between following two cases

1. a certain number (say *m*) of heavy wheel loads of magnitude P and
2. twice this number (2*m*) of smaller wheel loads of magnitude p/2.

For most wear-out phenomenon, including fatigue crack propagation and initiation the heavier wheel load of case (1) is more damaging to the rail compare to the larger number of lighter loads of case (2). Johnson and Besuner [16] proposed that the crack propagation rate *da/dN* is proportional approximately to the 4th power of stress leading to an effective usage parameter, *MGT-Effective* given by

$$\textit{MGT-Effective} = \left[\sum_{i=1}^{n} (MGT_i)^4\right]^{1/4} \qquad (1)$$

where n = total number of wheels passing through the curve section.

Kalousek and Magel [17] introduced the magic wear. It is the boundary between sufficient and insufficient wear – the optimal, or "magic", wear rate – represents a state of balance in which the development of fatigue is arrested by wear or metal removal. The magic wear rate is achieved when the surface material wears just enough to prevent surface fatigue cracks from propagating and causing fatigue failures. A combination of lubrication and light, but frequent, profile grinding is generally required to obtain the optimal condition. While every wheel/rail system has an optimal, or magic, wear rate, the rate is not constant. The optimal value changes with the hardness of the materials, the average contact stresses, the wheelset-steering performance, and coefficient of friction, or the effectiveness of the lubrication. Magel and Kalousek [18] recommended an interval of 8-12 MGT (280-300 BHN standard carbon rail) and 12-25 MGT (360-380 BHN premium rail) for sharp curves in most environments. It is important to scientifically determine grinding interval (how frequently the grinding should be conducted) and grinding depth (how much surface material should be removed). This knowledge is useful to improve the efficiency of grinding work and decrease in the maintenance cost. It is a big challenge for rail players to best utilize the available grinding power (and grinding budget) to maximise the effectiveness of the rail grinding program. Currently rail players around the world take decisions based on Visual Qualitative Checks (VQC), Non Destructive Testing (NDT), assumed Million Gross Tonnes (MGT) for axle loads and Traffic Density (TD). This is a slow, time consuming and costly process of decision making leading to a risky operating condition in between inspections. An ongoing cost effective data acquisition for condition monitoring of rail track would be able to overcome risks between periodic checks. Inspection is the key in reducing rail breaks, defects and catastrophic failures. There are three aspects to rail inspection, (1) technology used, (2) frequency of inspection, and (3) actions specified when a defect is detected. Inspection is generally based on line speed, track conditions, traffic tonnage and axle load and traffic type. Sawley and Reiff [19] show that broken rails are generally result from the following major factors:

- Poor inspection procedures. This may be due to poor operator training, out-of calibration equipment.
- Surface conditions interfering with the ultrasonic signal. It is known that surface damage, such as cracks, spalls, and flakes can hinder ultrasonic inspection by affecting transmission of ultrasound into the out of the rail. Figure 4 shows detection undetected fatigue cracks on rail surface.

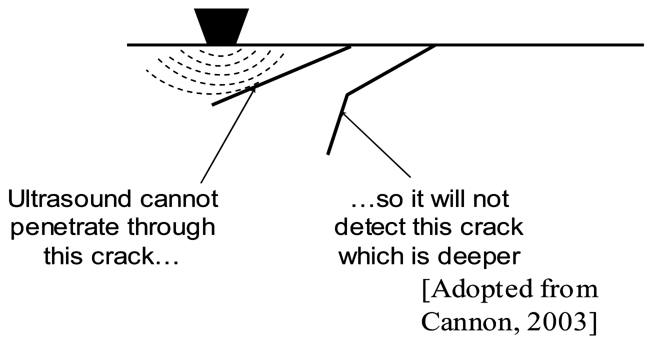


Figure 4: Detection undetected fatigue cracks on rail surface [20]

- Defects those are inherently difficult to detect using existing technologies. Ultrasonic signals enter the head of rail and, though they travel down the web section, they are not able to find defects in the outer edges of the foot. Cracks towards, the outer edges of the head are also difficult to locate, as are vertical/transverse crack such as those found in thermite welds.
- Inspection intervals that are longer than optimum. If inspection interval is too long, defects can grow from non-detectable to critical sizes between inspections.

Rails are regularly inspected by ultrasonic techniques using inspection vehicles with manual methods (including walking stick) used mainly for verification tests and tests at special areas such as switches and crossing. However, ultrasonic can only inspect the core of materials; that is, the method cannot check for surface and near-surface cracking where many of the faults are located. Because of this, some other techniques other than the ultrasonic method are used. Four main types inspection methods commonly used are [21]:

1. Visual inspection.
2. Dye penetrant inspection and magnetic particle inspection.
3. Eddy current testing.
4. Radiography

All of the above techniques could not able to detect surface and near-surface cracking defects accurately. In many cases an investigation will be performed to try and identify the cause of the failure. On many occasions the defect may be classified as undetectable at the time of test because it may have been too small or the surface condition of the rail may have presented additional ‘noise’ that may have masked the defect. Also, the cause of the broken rail may have been something such as a wheel flat. In these instances a latent defect likely to be found at the next test may become a catastrophe when subjected to the impact of a train wheel with flats. Cannon et al., [20] estimated annual vehicle inspections followed by manual verifications of detected defects, inspection costs are about €70 million per year for 0.5 million kilometres. This indicates the importance of the inspection technology, frequency interval and the detection of undetected defects. Therefore, there is need for detail study and analysis of inspection techniques to detect rail defects. Inspection before grinding, after grinding and inspection of lubricators, their performance, application and effectiveness of lubricant need to be studied for better understanding and for effective maintenance decisions to reduce operational risks and costs.

Rail replacement is most expensive operation in rail maintenance system. Rail infrastructure owners are trying to rectify the detected defects with various maintenance procedures. If the defect is at the initial stage then rectification could be postponed until it reaches to critical stage. If the rail defect is critical, immediate overhaul could be done to reduce risk of rail breaks and catastrophic failures. It is an expensive maintenance procedure even to rectify or to replace a small segment of rail defect during winter (below -25^0 Centigrade). This can be reduced only through accurate inspection frequency intervals during summer and in spring before the winter to reduce maintenance costs.

3.1 COST MODEL FOR RAIL MAINTENANCE DECISIONS

3.1.1 Modelling Lubrication and Wear rate

Wear rate *(I)* can be estimated as a linear function of $T\lambda/A$ parameter, where T is tangential force, λ is the relative slippage, and A is the Hertzian contact area.

$$I=T\lambda/A \tag{2}$$

Zakharov et al., (1998) [22] has conducted laboratory simulated tests of wheel flange and rail gauge face wear particularly for severe and catastrophic wear modes in unlubricated conditions and at a relatively stable coefficient of friction the specific volume wear rate *(I)* can be expressed as

$$I=kp\lambda^2 \tag{3}$$

where p is contact pressure at the contact patch and λ relative slippage varying from 0 to 1 and k is the coefficient of friction rail-wheel interaction. The relative slippage depends on the angle of wheel to rail attack and curve radius. It is important to evaluate the gauge face friction before positioning lubricators in order to determine the point of contact between rail-wheel interaction. Coefficient of friction is the classification of surface roughness which reduces wear on rail surface. The co-efficient of friction μ together with film parameter, Λ is shown in Figure 4. Coefficient of friction is expressed as:

$$\mu = f/w_z \tag{4}$$

Where f is tangential friction force and w_z is the normal applied load. It is found that film thickness parameter, Λ influence the co efficient of the friction, μ. The pressure builds up in moving surface separate the asperities from the boundary lubrication to the mixed film lubrication. Lubrication cost can be estimated assuming cost of lubricant (C_l) selected for application of particular curve section *(r)* of rail segment *(s)* under known weather and environmental conditions. Let C_{mirs} be the cost of maintenance of each lubricator. It includes lubricator servicing, checking and filling of lubricator tanks, evaluating the performance of lubricators and checking the blades and plungers for each lubricator.

C'_{mirs} = cost of emergency repair during the failure of i^{th} lubricator or lubricant leak or spilling of lubricant.

C_{pirs} = cost of the personnel involved in maintenance of i^{th} lubricator.

Differential Wear loss (W_{irs}) = Total rail wear before lubrication – Loss of material after lubrication for i^{th} lubricator of curve section r of rail segment s.

Where i is the index.

C_r = *Cost of rail material per meter per kg*

Therefore, the total cost of differential wear loss for particular curve section *(r)* of rail segment *(s)* can be expressed as

$$TC_{W_{irs}} = W_{irs} * C_r \tag{5}$$

Total cost of lubrication for i^{th} lubricator at curve section r and rail segment *(s)* can be expressed as

$$TC_{l_{irs}} = C_L + C_{mirs} + C'_{mirs} + C_{pirs} \tag{6}$$

If

$$TC_{l_{irs}} \leq TC_{W_{irs}} \text{ then the lubrication is effective} \tag{7}$$

This is individual comparison of total cost of each lubricator and total wear cost for particular curve section (r) and rail segment *(s).* Some places it is difficult to compare these costs due to different curve radius and rail size, age and condition of rail. The results will be more appropriate if we consider average total cost of n number of lubricators for curve section for a rail segment. Therefore, the total average cost of n number of lubricators for curve section (r) and rail segment (s) can be expressed as:

$$TC_{nlrs} = \sum_{i=1}^{n} TC_{l_{irs}} \tag{8}$$

The total average wear cost for n lubricators of curve section (r) and rail segment (s) can be expressed as:

$$TC_{wnrs} = \sum_{i=1}^{n} TC_{W_{irs}} \tag{9}$$

The ratio of Equation (8) and (9) can be used to determine the percentage of average costs of wear and lubrication for particular curve section at a rail segment under consideration. Therefore, the ratio can be expressed as,

$$\text{\% of average costs} = \frac{TC_{nlrs}}{TC_{wnrs}} \times 100 \tag{10}$$

The average costs can be compared to determine the effectiveness of lubrication and differential wear costs. The ratio of these costs would be used to determine the percentage of savings in terms of lubrication costs and also for evaluation of lubricator's performance. Then the total cost of maintaining n lubricators for particular curve section (r_i) of rail segment (s_i) per year

$$T_c = \sum_{i=1}^{n} T_{Cl_{irs}} \tag{11}$$

3.1.2 Modelling Rail grinding cost

Let G be the cost of grinding per pass per meter and n_i be the number of grinding pass for i^{th} grinding, L be the length of rail segments (0-300, 300-450, 450-600, 600-800 meters of curve radius sections) under consideration, N be the total number of periods up to safety limit for renewal, and r be the discounting rate per period. It is assumed that payments are made to subcontractors after each of the (N-1) grinding.

Therefore the annuity cost for rail grinding is given by:

$$c_g = \{\sum_{i=1}^{N-1} (G * n_i * L)/(1+r)^i\} * r_y /(1-(1/(1+r_y)^Y)) \quad (12)$$

3.1.3 Modelling Rectification and Replacement costs

Once rail defect is detected it is important to decide what type of maintenance activity is required to reduce risk of rail breaks and derailments. Currently rail infrastructure owners are following three types of maintenance activities. They are

- Minimal repair
- Overhaul and
- Replacement (for details see [23])

Let c_{re} be the expected cost of replacement for segment L and consists of labour, material, and equipment, consumable and down time cost for rail replacement. Let I be the cost of current investment in new rail. In this model the cost of replacement is assumed to be occurring at the beginning of each year and is simplifies as the annual spread over of investment of new rail. Then c_{re} is given by:

$$c_{re} = I * (1-(1/(1+r)))/(1-(1/(1+r_y)^y)) \quad (13)$$

where r is discount rate.

3.2 Modelling Operational Risks due to Wear-Fatigue-Lubrication Interaction

Let cost per rectification of rail breaks on emergency basis, C_r be modelled through $G(c)$, and is given by

$$G(c) = P[C_r \le c] \quad (14)$$

For an example, if $G(c)$ follows exponential distribution [24], then it is given by

$$G(c) = 1 - e^{-\rho c} \quad (15)$$

where $\bar{c}$ denote the expected cost of each rail break repair on emergency basis and is given by:

$$\bar{c} = [1/\rho] \quad (16)$$

Let k be the expected cost of repairing potential rail breaks based on NDT in a planned way and a be the expected cost per derailment. Then k and a could be modelled in similar manner. The risk cost associated with rail break and derailment is based on the probability of NDT detecting potential rail breaks, rail breaks not detected by NDT, derailments and associated costs. It also depends on lubrication, grinding, and rectification and replacement maintenance activities of rail. Let $P_i(A, Fatigue, s)$ be the probability of undetected potential rail breaks leading to derailments based on rail head area, RCF, speed of train. Let $P_i(B)$ be the probability of detecting potential rail breaks based on rail head area, fatigue and speed of train. Rolling contact fatigue is given by Million Gross Tonnes (MGT). Railhead area is determined by wear and preventive rail grinding based on MGT [25]. When expected number of failures are modelled as Non Homogeneous Poisson process and is given by $E[N(M_i+1, M_i)$ then the risk cost is given by:

$$C_r = \left\{ \sum_{i=0}^{N} \frac{E[N(M_{i+1}, M_i)] * [P_i(B) * k + (1 - P_i(B)) * (P_i(A, Fatigues)) * a + (1 - P_i(A, Fatigues)) * \overline{C}]}{r/(1-(1/(1+r)^N))} \right\} \quad (17)$$

Where r is discount rate.

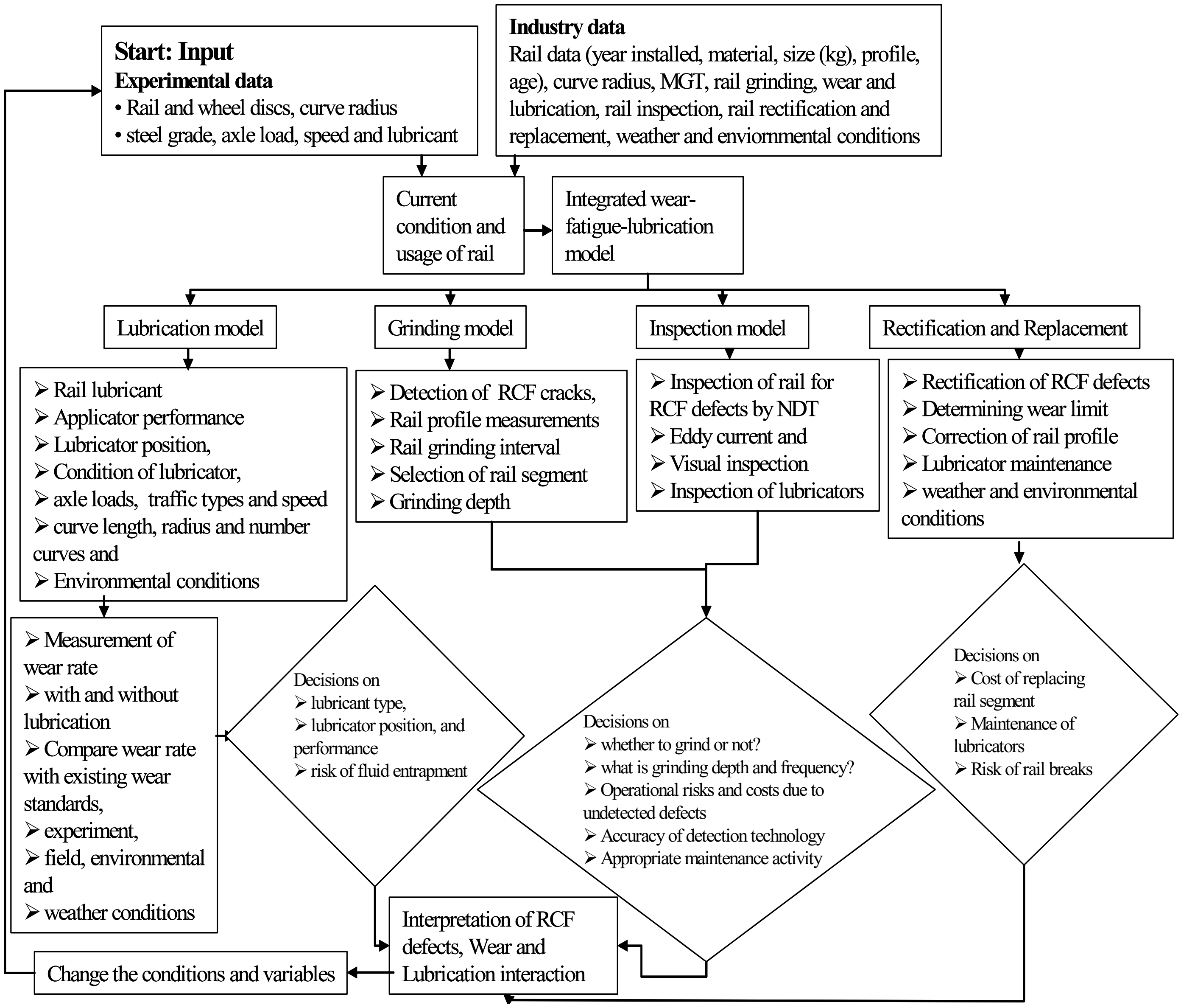


Figure 5: Proposed integrated model for wear-fatigue-lubrication and grinding

Figure 5 shows proposed model for wear-fatigue-lubrication interaction, grinding and inspection maintenance activities. The model can be used for effective maintenance decisions on lubricators, type of lubricant, effectiveness of lubricant, grinding death and interval, inspection techniques accuracy and interval, defect rectification and replacement of rails to minimise operational risks and costs.

4 SUMMARY AND CONCLUSION

In this paper an integrated model proposed for wear-fatigue-lubrication interaction. It also covers grinding and inspection activities. The proposed model would be useful for analysing wear-fatigue-lubrication interaction considering grinding, lubrication, inspection, rectification and replacement decisions. Cost models proposed here could be used for rail-wheel maintenance and replacements decisions. It is able to evaluate performance of lubricators and grinding and inspection. Authors are currently working in this area and results will be published in the near future.

5 REFERENCES

[1] Nilsson, R., (2005) "On wear in rolling/sliding contacts", Doctoral thesis, Department of Machine Design, Royal Institute of Technology, SE-100 44 Stockholm, TRITA – MMK 2005:03 ISSN 1400-1179, ISRN/KTH/MMK/R-05/03-SE.

[2] Franklin, F.J., Weedab, G.-J., Kapoor, A., Hiensch, E.J.M., (2005) "Rolling contact fatigue and wear behaviour of the infrastar two-material rail", *Wear* 258, 1048-1054.

[3] Pandey, L. M., Saxena, D., Sayanna, E., Mookken, R. T., Srivastava, S. P., and Bhatnagar, A. K., (2000) "Railway Wheel Flange Lubrication A New Approach", Indian Oil Corporation Limited, R & D Centre, Sector 13, Faridabad.

[4] IHHA, (2001) "Guidelines To Best Practices For Heavy Haul Railway Operations: Wheel and Rail Interface Issues", *International Heavy Haul Association*, 2808 Forest Hill Court, Virginia Beach, Virginia 23454, USA.

[5] Larke R, (2003) "Survey Of Wheel / Rail Lubrication Practices", Research and Development Programme, Rail Safety Standard Board, UK, Report no: ITLR-T12442-001.

[6] Sid D, Wolf, E (2002), "Transportation FOR THE 21ST CENTURY", Trac Glide Top-of-Rail Lubrication System, Report from Department of Energy, USA.

[7] Sroba, P., Magel, E., Prahl, F., (2003) "Getting the most from rail grinding: the tiny interface where steel wheel meets steel rail can maintain its smooth, efficient function much better with preventive grinding practices", *Railway Track and Structures*, Dec 2003.

[8] Cannon, D. F. and Pradier, H., (1996) "Rail rolling contact fatigue – Research by the European Rail Research Institute", *Wear*, 191, 1-13.

[9] Ringsberg, J.W., (2001) "Life prediction of rolling contact fatigue crack initiation", *International Journal of Fatigue*, Vol. 23, pp. 575-586.

[10] Sawley K., Kristan J., (2003) "Development of bainitic rail steels with potential resistance to rolling contact fatigue", *Fatigue & Fracture of Engineering Materials and Structures* 26, 1019-1029.

[11] The Guardian, October 7, 2005 http://www.guardian.co.uk/hatfieldtraincrash/story/0,,1587278,00.html

[12] Fletcher, D. I. & Beynon, J. H., (2000) "The effect of contact load reduction on the fatigue life of pearlitic rail steel in lubricated rolling-sliding contact. *Fatigue & Fracture of Engineering Materials & Structures* 23.8, 639-650.

[13] Railtrack plc., (2001) "Rolling Contact Fatigue in Rails: A Guide to Current Understanding and Practice", RT/PWG/001, Issue 1, February 2001.

[14] Hamrock, B.J. (1994). Fundamentals of Fluid Film Lubrication. McGraw Hill Book Co.

[15] Besuner, P. M., Stone, D. H., Schoenberg, K. W., and de Haretta, (1978) "Probability Analysis of Rail Defect Data", *Proc. Heavy Haul Railways Conf.*, Perth, IE(Aust.).

[16] Johnson, D. P., and Besuner, P. M., (1977) "Engineering Cost Risk Analysis of Defective Rail", *Association of American Railroads Report*, R-2, January, 1977.

[17] Kalousek, J. and Magel, E., (1997) "Achieving a balance: The "magic" wear rate", *Railway Track & Structures*, 50-52.

[18] Magel, E., Kalousek, J., (2002) "The application of contact mechanics to rail profile design and rail grinding", "*Wear*", Vol. 253, Issues 1-2, pp. 308-316.

[19] Sawley, K. & Reiff, R. (2000) "An assessment of Rail track's methods for managing broken and defective rails." Office of the Rail Regulator, Transportation Technology Center Inc., Association of American Railroads. Pueblo. Colorado. A2 – A14.

[20] Cannon, D.F., Edel, K.O., Grassie, S. L. & Sawley, K. (2003) "Rail Defects: an overview", *Fatigue & Fracture of Engineering Materials & Structures*, 26.10, 865-887.

[21] Cope, G. H., (1993) "British Railway Track", Design, Construction and Maintenance.

[22] Zakharov. S., Komarovsky. I., Zharov. I., (1998) "Wheel Flange/Rail Head Simulation", *Wear*, 215.18-24.

[23] Reddy V (2004) "Modelling and Analysis of Rail Grinding & Lubrication Strategies for Controlling Rolling Contact Fatigue (RCF) and Rail wear", Australian Digital Thesis (ADT) Program (Research Thesis).

[24] Crowder, M. J., Kimber, A. C., Smith, R. L. and Sweeting, T. J., (1995) "Statistical Analysis of Reliability Data", Chapman & Hall, London SE1 8HN, UK.

[25] Chattopadhyay, G., Reddy, V. and Larsson, P. O., (2003) "Integrated Model for Assessment of Risks in Rail Tracks under Various Operating Conditions", *International Journal of Reliability and Applications*, 4.4, 183-190.

DETECTION AND ANALYSIS OF DETERIORATION TRENDS IN CONSTRUCTION MACHINES OPERATION

Yuhiko Kiyota [a] Gancho Vachkov [a] Koji Komatsu [a] Satoshi Fujii [b] and Nobuyuki Kimura [c]

[a] *Dept of Reliability-based Information Systems Engineering, Faculty of Engineering, Kagawa University, Hayashi-cho 2217-20, Takamatsu City, Kagawa 761-0396, Japan; E-mail: {kiyota, vachkov, komatsu}@eng.kagawa-u.ac.jp*

[b] *Hydraulic Excavator Development Center, Shin Caterpillar Mitsubishi LTD., Akashi-Shi, Hyogo-ken 674-8686, Japan.*

[c] *Shikokukiki Corp., Kankou-doori 2-2-15, Takamatsu-City, Kagawa 760-8518, Japan*

Abstract: The paper proposes a computational strategy for discovering trends of changes and gradual performance deterioration of construction machines and other complex systems. If a significant change in the machine performance is detected, it further activates a fault diagnosis procedure in order to find the possible faulty condition. In order to send the information from various sensors of the construction machine to the maintenance center in an efficient form, a special information compression method is proposed in the paper. It uses original unsupervised learning algorithm to locate the given number of neurons from the parameter space in the densest data areas. These neurons are considered as information granules, which are later on sent in a wireless way to the maintenance center to represent the current machine operation. Two information recovery procedures are also proposed in the paper for analyzing the compressed information from the neurons. The first one is a specialized version of the moving window method, while the other is fuzzy inference-based approach for discovering possible machine deterioration. Results based on real experimental data from a hydraulic excavator are used to explain the proposed computational strategy and its merits.

Key Words: Information Compression, Information Recovery, Unsupervised Learning, Neurons, Fuzzy Inference, Maintenance, Deterioration, Information Granules

1 INTRODUCTION

Construction machines, such as excavators, wheel loaders and others are complex systems that use many different sensors to provide adequate information for their control, monitoring, fault diagnosis and maintenance. As a result, a big number of data, taken by regular sampling over the daily operation of the machine are collected on-board and should be dispatched to the Maintenance Center for evaluation and analysis of a possible deterioration or abnormality in the machine operation.

The real problem in such analysis and detection of changes and deterioration trends is the high dimensionality of the data space and the large number of the obtained data, which makes any direct human decision quite subjective and ambiguous. The first and the simplest question to be answered, when looking at the multidimensional collected data set is whether the current machine operation condition is similar or differs significantly from the standard “normal operation” mode. In other words detecting the changes in system performance in automatic way is an important task which, if properly solved can provide a basis for solution of the next fault diagnosis problem.

There is a big interest in the literature on the problems related to the fault detection, diagnosis and performance evaluation in industrial and other complex systems, as shown in [1-5]. The widely used methods here involve different modifications of neural models [3], fuzzy models [4] or knowledge-based clustering [5]. The authors interest in this area and some of their latest results can be also found in [6,7].

In this paper we propose an efficient computational strategy for automated analysis and detection of changes and deterioration trends in operation of construction machines. It is based on the concept of information compression of the original large number of “raw data” into a smaller portion of “information granules” [5] in the form of neurons in the multidimensional space. Thus a very economical information transfer between the construction machine and the maintenance center is achieved. Finally, two additional methods are proposed as tools for recovering and further analysis of the information.

The following Fig. 1. is a graphical illustration of how complex and ambiguous is the problem for detection of similarities (or differences) between various operations of a construction machine. It displays four different operations of a hydraulic excavator. By just observing visually the *raw normalized* data in Fig. 1., it would be difficult for the human operator to properly judge which operations are “similar”, as well as which of them are “normal” and which belong to “abnormal” operation of the machine. All the human operator can do is to compare the shapes of the “data clouds” in the respective pairs of two-dimensional spaces.

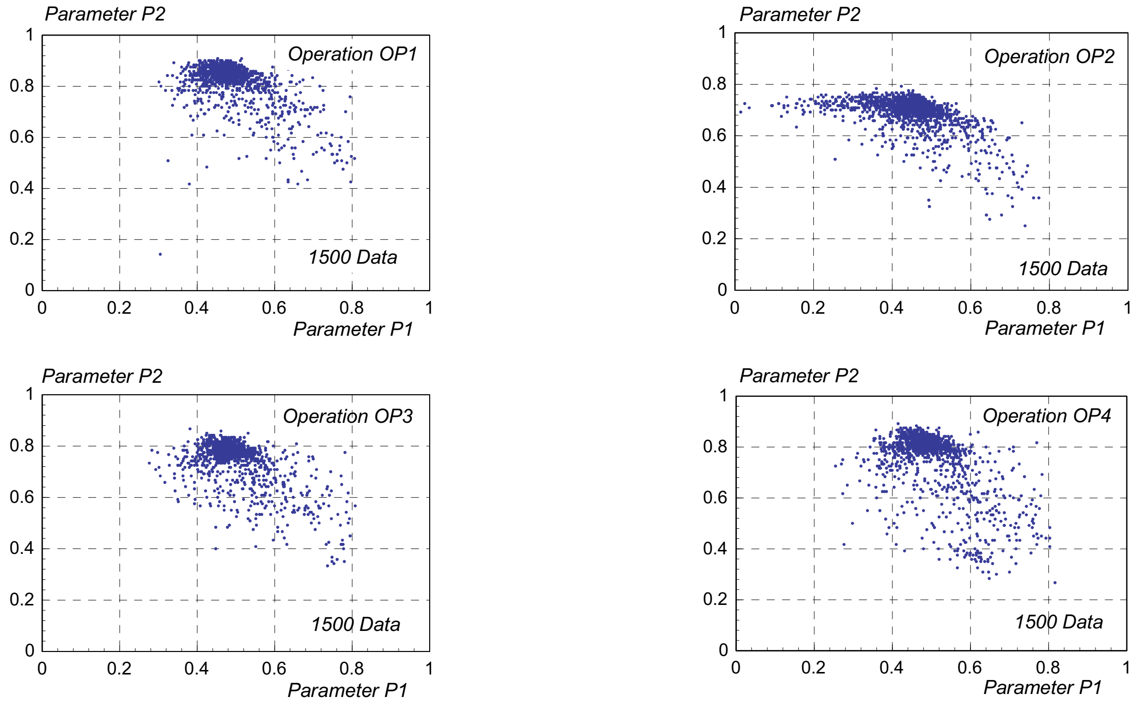


Figure 1. Example of four different operations of a construction machine, shown as plots of normalized “raw data”.

The remaining part of this paper is organized as follows. The proposed system for information compression and information recovery is explained in the next section *2*. Section *3* deals with the method for information compression, while section *4* presents two methods for information recovery. In section *5* different (normal and abnormal) operations of a hydraulic excavator are analysed by the proposed methods and system. Finally section *6* concludes the paper.

2 THE PROPOSED IC-IR SYSTEM FOR INFORMATION COMPRESION AND INFORMATION RECOVERY

Our basic assumption is that the *Operating Site* (OS) of the construction machine and the *Maintenance Center* (MC) are remotely located at a big distance from each other. Therefore an economically efficient system for transmitting the most relevant information for the current (daily) operation of the machine to the maintenance center is needed. This information is then processed and analysed in the MC in order to make an appropriate conclusion about the current state (normal, abnormal, faulty) of the machine and to take appropriate measures for correction (repair), if needed.

Since many thousands of data from different parameters are usually measured by the respective sensors during the daily operation of the machine, it s not practical to send all these “raw data” directly the maintenance center (MC) for a further analysis. Here there could be at least two different approaches to alleviate this high dimensionality problem.

One approach is to make an “on-board” pre-processing (simplified analysis) of the data and then to send the results of this analysis in a wireless way to the MC for final decision making. The merit of this method is that it is economical (a small number of data are transmitted between OS and MC), but at the same time it has two demerits, as follows. The first one is that the usually small “on-board” computing power puts limitations on the complexity of the analysis, which results in a simple (not sufficient) data analysis being done “on-board”. The second demerit is that in this way we are losing “on the spot” a big portion of information which cannot be further recovered at the MC.

Another approach, which we propose in this paper, is to make a kind of "information compression" (IC) based on the all measured raw data "on-board" and to transmit this compressed (highly reduced) information to the MC in a wireless way. Then, at the MC a respective "information recovery" (IR) procedure is run which is followed by information analysis and appropriate decision making step. In such way, the current daily operation of the construction machine can be analysed, evaluated, classified (as normal, deteriorated or faulty) at the maintenance center, as well as to be compared to other operations of the same machine in different (past) time periods.

The proposed system, based on the IC-IR approach for analysis of the deterioration trends has two clear merits. The first one is that the compressed information "on-board" is recoverable (to a high extend) at the MC. This means that we are not losing a lot of useful information, as in the previous mentioned approach. The second merit is that the compressed information can be efficiently transmitted to the MC, which makes the whole hardware system, based on this approach very practical.

The block-diagram of the IC-IR system for information compression and information recovery, based on the above proposed approach is shown in Fig. 2.

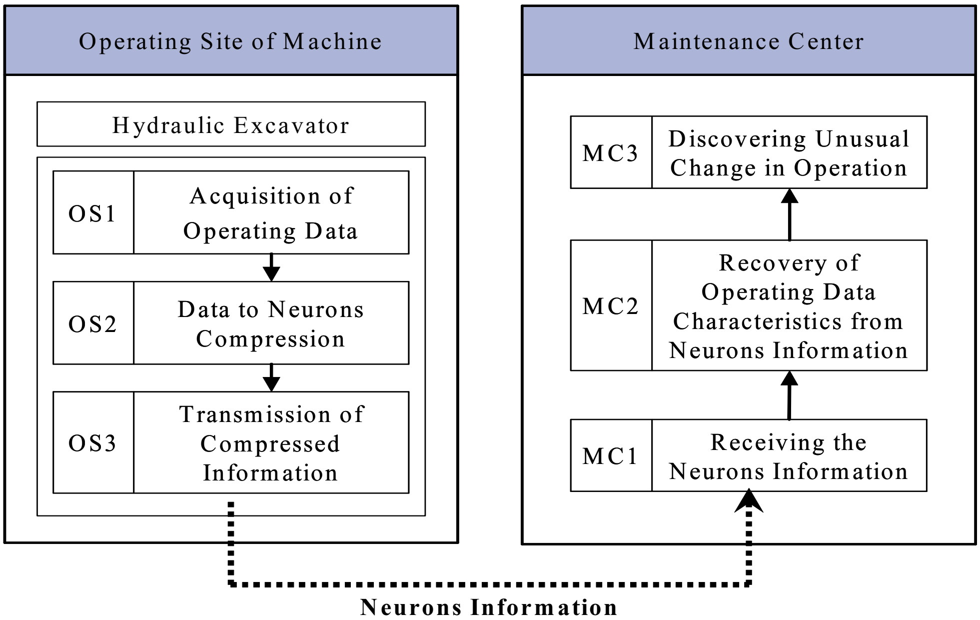


Figure 2. The Proposed IC-IR System for Information Compression and Information Recovery.

As seen from the figure, there are three main steps to be performed at the Operation Site of the construction machine, denoted as OS1, OS2 and OS3 and three other computation steps at the Maintenance Center, denoted as MC1, MC2 and MC3.

Step OS1 is dealing with the real-time data measuring of a fixed number of parameters and data acquisition. Normally the data are measured with a sampling period of *1 sec* in order to represent the dynamical behavior of the machine during its operation. As a result, several thousands of data are accumulated in the "on-borad" computer memory.

Step OS2 is the essential part of the whole IC-IR system, because it performs the information compression, which is explained in details in the next Section of the paper. In short, this step creates a kind of "information granules" in the form of preliminary fixed number of neurons, appropriately trained in the parameter space, so that to best represent the data structure of the "raw data".

Step OS3 is a wireless transmission of the compressed information from the OS to the MC, i.e. transmission of the information granules (neurons) with all the related information about their parameters.

Once the compressed information is transferred and received at the Maintenance Center (step MC1 in Fig. 2.), then step MC2 performs some computational methods and algorithms for information recovery. In Section 4 of the paper we suggest and explain two methods and algorithms, which represent two different aspects of the information recovery. The first one is based on the window moving average and the second one is an original fuzzy inference method for analysis of the relationship between the parameters.

The results from the information recovery at step MC2 are further on used in the last step MC3, which is a final decision making procedure for confirmation (or rejection) of a possible deterioration (or faulty) state of the machine. This step includes also a possible automatic fault diagnosis, which is not a research subject in this paper.

The whole proposed IC-IR system in Fig. 2. is currently in a status of implementation for performance analysis and evaluation of the operations of hydraulic excavators, including detection of abnormal operations.

3 THE METHOD FOR INFORMATION COMPRESSION

Further on we suppose that a big number (usually many thousands) of M "raw data" $\mathbf{x}_s, s=1,2,...,M$ have been collected at the Operating Site during the current operation of a construction machine. Each data point $\mathbf{x}_s$ represents the measurements from K parameters, which are dynamically changing during the operation, i.e. $\mathbf{x}_s = [x_{s1}, x_{s2}, ..., x_{sK}]$, $s = 1,2,...,M$.

The general concept of the information compression method proposed in this paper is to replace the complete data set of M data with a much smaller set of N ($N \quad M$) *information granules* that carry as much as possible of the original amount of information given by the "raw data".

The most important part of the information that should not be lost in the process of information compression is the original *data structure* for the current operation of the machine. This structure is best represented by the overall *data shape* (in the form of "cloud" in the K-dimensional space); as well by the *data density distribution* in the different areas of the data "cloud". These two characteristics of the data play important role in the final procedure (Step MC3 in Fig. 2.) for detection of deterioration trends in the machine performance as well as for fault diagnosis purposes.

Further on we consider all N information granules obtained by the information compression method as a kind of *neurons* in the K-dimensional parameter space, which carry the following three types of information:

Neuron *centers*: $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_N\}$; Neuron *weights*: $\mathbf{G} = \{G_1, G_2, ..., G_N\}$ and Neuron w*idths*: $\mathbf{W} = \{\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_N\}$. Please note that the centers in $\mathbf{C}$ and the widths in $\mathbf{W}$ are K-dimensional vectors, while the weights in $\mathbf{G}$ are scalars.

Neuron *centers* $\mathbf{c}_i = [c_{i1}, c_{i2}, ..., c_{iK}]$, $i = 1,2,...,N$ show the place (position) of each neuron in the K-dimensional space: Generally neurons should be surrounded by as much as possible "raw data" points, in order to represent (compress) the information in a suitable way. In other words these centers serve as prototypes of local clusters in the space [5,8].

Neuron *weights*: G_i, $i = 1,2,...,N$ can be viewed as *strengths* of the neurons, in a sense of number of the data points that are represented (managed) by each neuron.

Finally, neuron *widths* $\mathbf{w}_i = [w_{i1}, w_{i2}, ..., w_{iK}]$, $i = 1,2,...,N$ are a kind of measure for the neuron *activities,* which indicate *how large* is the area around each neuron center for which it is "responsible".

The proposed information compression method first locates the neuron centers by appropriate unsupervised learning algorithm and then defines the other two parameters, namely the neuron weights and neuron widths. The computational details and illustrations are given in the next sub-sections.

3.1 Location of the Neuron Centers by Unsupervised Learning.

The main goal of this computational procedure is to train the centers of the neurons in the most appropriate way, so that to locate all the neurons at the "best" places in the K-dimensional parameter space. A natural way to do this is to use the information about the density distribution of all available data points $\mathbf{x}_s, s=1,2,...,M$ in the space.

Basically, the learning methods in this case belong to the group of the competitive unsupervised learning methods and algorithms [8,9]. Among them are the well known and widely used clustering algorithms [5,8], the Self-Organized (*Kohonen*) Maps [9], Neural Gas Algorithms [10] etc. Because of their unsupervised nature, these methods are considered as pure "data-driven learning" (DDL) methods which tend to locate the neurons in the densest (the most "crowded") area of the input space.

We use in this paper the well known and popular Neural-Gas unsupervised learning algorithm, first reported in [10], with some minor algorithmic improvements and simplifications, which make it more suitable for our problem.

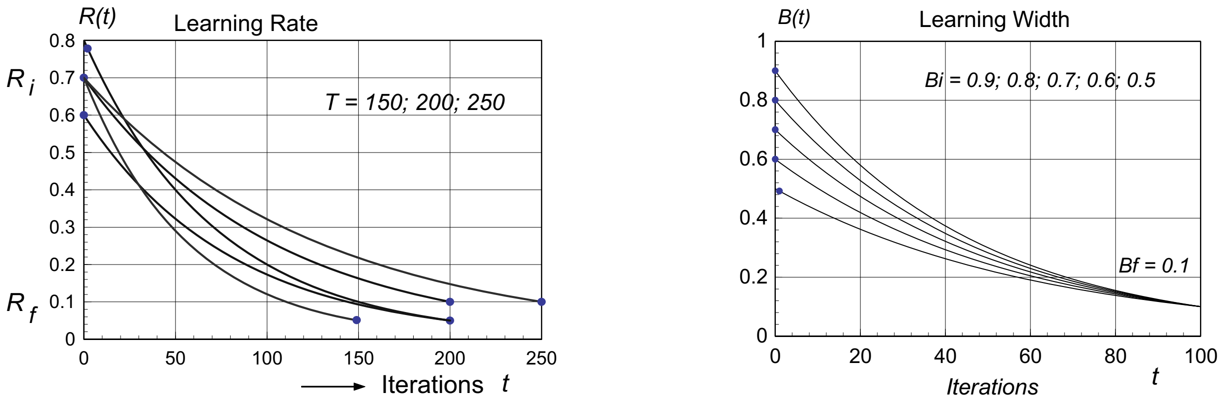
Initialization Step: There are some parameters that have to be set in advance, before the unsupervised learning starts. First of all a preliminary fixed number of N neurons ($N \quad M$) is assumed and their initial locations (centers) are set randomly in the K-dimensional space as follows: $\mathbf{c}_i^{\,0} = [c_{i1}^{\,0}, c_{i2}^{\,0}, ..., c_{iK}^{\,0}]$, $i = 1,2,...,N$. The maximum number of iterations T for the unsupervised learning is also set in advance and usually varies between several hundreds and several thousands. Another tuning parameter is the *learning rate* $R(t)$, $0 \le R(t) \le 1$, $t = 0,1,2,...,T$ which controls the speed and convergence of the learning process. It is defined by two initial parameters: R_i and R_f representing the initial and final learning rate, respectively. The learning rate depends also on the current number of iteration, as shown by the following exponentially decreasing function:

$$R(t) = R_i (R_f / R_i)^{t/T}, \; t = 0,1,2,...,T \tag{1}$$

Another tuning parameter in the unsupervised learning is the *learning width* $B(t)$ which controls the size of the "neighbourhood" of the neurons (explained later). In a similar way as in $R(t)$, the learning width $B(t)$ is defined by two initial parameters: B_i and B_f that represent the *initial* and *final* width respectively. It also decreases exponentially with the number of iterations, as follows:

$$B(t) = B_i (B_f / B_i)^{t/T}, \ t = 0,1,2,...,T \tag{2}$$

The following Fig. 3a) and Fig.3b) is an illustration of the learning rate and learning width evolutions during the iterations under different initial and final settings.



a)Exponentially Decreasing Learning Rate b)Exponentially Decreasing Learning Width

Figure 3. Evolution of *Learning Rate* and *Learning Width* under different initial and final settings.

<u>Learning Step:</u> This step is repeated for each iteration $t = 1,2,...,T$ and for all available data $\mathbf{x}_s$, $s = 1,2,...,M$. The purpose is to gradually move the centers $\mathbf{c}_i, i = 1,2,...,N$ of all N neurons into the areas in the parameter space with a bigger data density. This is done by the following general incremental learning rule:

$$\mathbf{c}_i(t) = \mathbf{c}_i(t-1) + \Delta\mathbf{c}_i(t), \ i = 1,2,...,N. \tag{3}$$

The increment $\Delta\mathbf{c}_i(t)$ for the movement of each neuron's center is computed as:

$$\Delta\mathbf{c}_i(t) = R(t) H_s(t,r) \left[\mathbf{x}_s - \mathbf{c}_i(t-1) \right], \ i = 1,2,...,N. \tag{4}$$

It is seen from (4) that the amount of learning (movement) of the *i*-th neuron in the *K*-dimensional parameter space depends on the following three factors:

- the current learning rate $R(t)$, computed by (1);
- the current *neighborhood* of the *i*-th neuron in respect to the *s*-th data point $\mathbf{x}_s$, computed by the *neighborhood function* $H_s(t,r)$;
- the current distance between the *s*-th data point and the *i*-th neuron (taken separately for each of all *K* parameters).

The *neighborhood function* $H_s(t,r)$ is an exponentially decreasing function, which is important and specific part of the Neural-Gas unsupervised learning algorithm. It can be viewed as a variable scaling factor for the *amount* of the update for each neuron, which depends on the relative closeness of the *i*-th neuron to the current s-th data point $\mathbf{x}_s$. This relative closeness is taken as the *ranking position* $r = 1,2,3,...$ of this neuron compared with all other neurons in the space. Finally the neighborhood function is computed as:

$$H_s(t,r) = \exp\left(-\frac{(r-1)^2}{B(t)} \right) \tag{5}$$

where $B(t)$ is the current learning width of the neighborhood function, computed by (2).

For a given data point $\mathbf{x}_s$, the ranking positions $r = 1, 2, 3, \dots, N$ for all N neurons are computed through a sorting procedure in ascending order of the *Euclidean distances* $D_S(n)$ between all the neurons n, $n = 1, 2, \dots, N$ and this data point:

$$D_S(n) = \sqrt{\sum_{j=1}^{K} \left(x_{s\,j} - c_{n\,j} \right)^2} \tag{6}$$

The neuron with the minimum distance to $\mathbf{x}_s$ (the closest neuron to this data point) is called "winning neuron" and gets the ranking position $r = 1$. The second closest neuron to the same data point gets position $r = 2$ and so on until certain neuron is ranked as the last one ($r = N$) being the furthest neuron to the data point $\mathbf{x}_s$.

Figure 4. shows the evolution of the neighborhood function during a learning with *T = 100* iterations. Since the learning width $B(t)$ in (5) is a monotonically decreasing function of the iterations, the neighborhood function (5) is gradually becoming narrower, which results in a smaller amount of update for a neuron with the same ranking position r. From (5) it is clear that the "winning neuron" gets always the full amount of update (4) that is $H_s(t,r) = 1.0$ for $r = 1$. Other neurons get smaller updates, according to their ranking, as seen from Fig. 4.

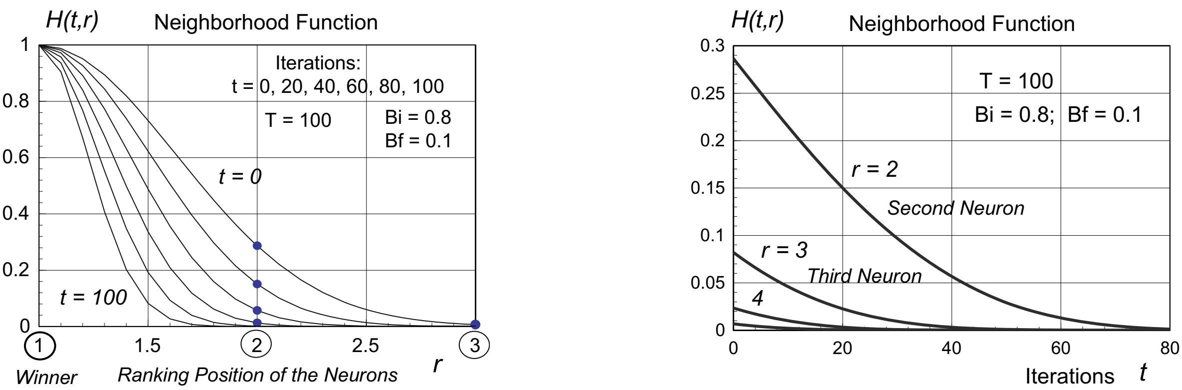
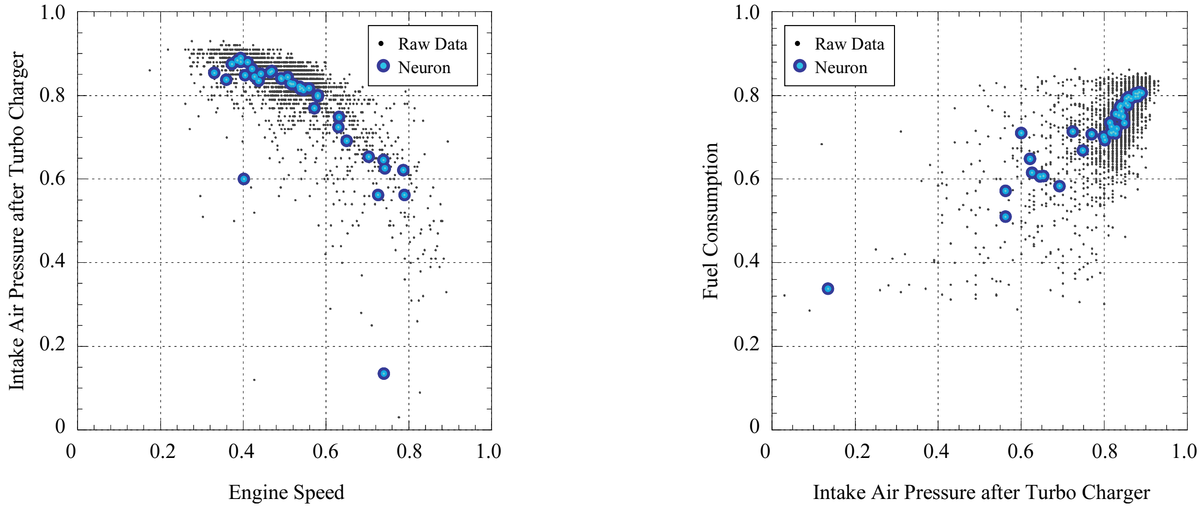


Figure 4. Evolution of the Neighbourhood Function during iterations and updates of the neurons.

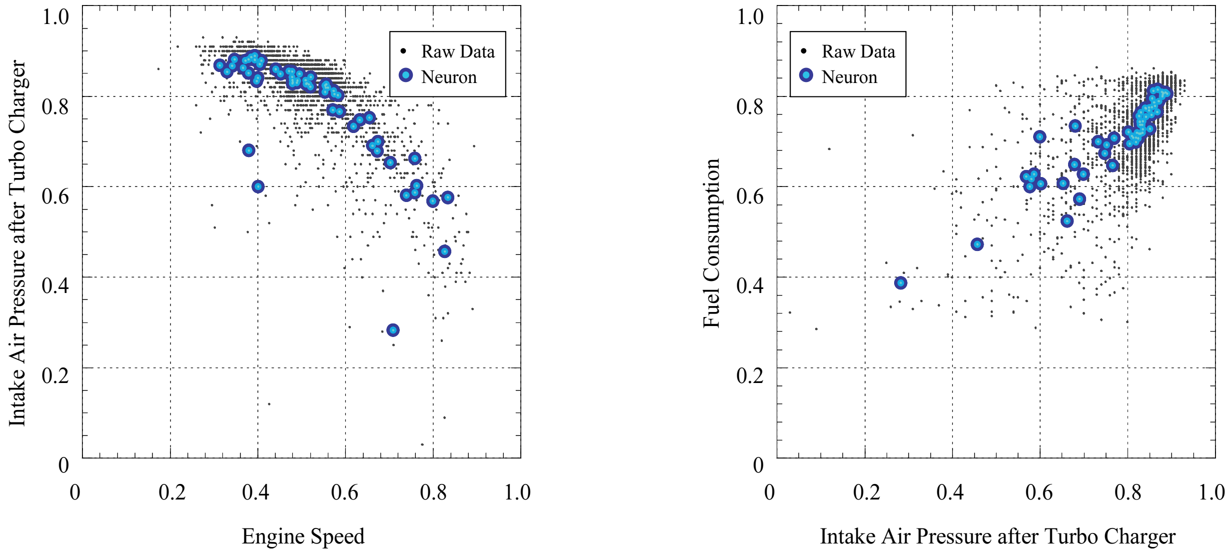
3.2 Illustration of the Information Compression by the Unsupervised Neural-Gas Learning

The next two figures, Fig. 5. and Fig. 6. serve as illustration of how the Neural-Gas unsupervised learning algorithm compresses the information carried by a set of *2515* raw data with three different parameters (*K = 3*), taken from a hydraulic excavator with a sampling period of *1 sec*. Figure 5 shows the case of compression by using a set of *40* neurons (*N = 40*), while Figure 6 is a compression of the same data set by using *N = 60* neurons. In both cases the learning is done with the following learning parameters: $T = 300$; $R_i = 0.5$; $R_f = 0.05$; $B_i = 0.7$; $B_f = 0.1$. It is visually understood that even if there is not a big difference between the two figures, the bigger number of neurons in Fig. 6. allows a better preservation of the structure (shape) of the original raw data set. Therefore the neuron set from Fig. 6.can serve as a set of information granules for transmitting the compressed information to the Maintenance Center for further recovery and information analysis. The problem of a proper selection of the number of neurons for a minimum loss of information is empirically solved in this paper, but could be done in more rigorous way by setting an appropriate criterion.



(a) (b)

Figure 5. Information Compression of $M = 2515$ raw data from a normal operation, by using $N = 40$ neurons.



(a) (b)

Figure 6. Information Compression of $M = 2515$ raw data from normal operation, by using $N = 60$ neurons.

3.3 Computing the Neuron Weights $\mathbf{G}$

As mentioned above in the section, the neuron weights $\mathbf{G}=\{G_1, G_2, ..., G_N\}$ are scalar values, which serve as *strength levels* for each neuron, evaluated by the number of data points that are represented by (or "belong" to) this neuron. In order to find the number of data, "belonging" to each neuron we use again the notion of "wining neuron" from the Neural-Gas learning algorithm.

First of all, for each data point $\mathbf{x}_s$, $s=1,2,...,M$ we find the respective "winning neuron" n_s, $1 \le n_s \le N$ as the closest neuron to this data, i.e. the neuron with a minimum Euclidean distance (4) to $\mathbf{x}_s$:

$$D_s(n_s) = \min_{1 \le n \le N} \{D_s(n)\} \tag{7}$$

Then this information is used for splitting all M data into N subsets: $\mathbf{M}_n$, $n=1,2,...,N$, each of them containing the data points m_n, $n=1,2,...,N$ for which the same n-th neuron has been "winning neuron". Speaking more precisely, the following conditions hold:

$$m_n = |\mathbf{M}_n|;\ 1 \le m_n \le M,\ n=1,2,...,N;\quad \sum_{n=1}^{N} m_n = M = |\mathbf{M}|,\ \text{where}\ \ \mathbf{M} = \bigcup_{n=1}^{N} \mathbf{M}_n \tag{8}$$

The normalized weights of the neurons are computed simply as follows:

$$G_n = m_n / M;\ 0 < G_n \le 1,\ n=1,2,...,N,\ \ \text{with}\ \ \sum_{n=1}^{N} G_n = 1. \tag{9}$$

It is worth to note that each subset of data $\mathbf{M}_n, n=1,2,...,N$ forms the so called "*Voronoi polygon"* [9,10] in the K-dimensional space.

Illustration of a *rough* information compression of a set with $M = 759$ data by a set of $N = 6$ neurons, trained with the Neural-Gas algorithm, is presented in the following Fig. 7. It shows also the shape of all 6 "*Voronoi polygons*" and their respective number of data points m_n, $n=1,2,...,6$.

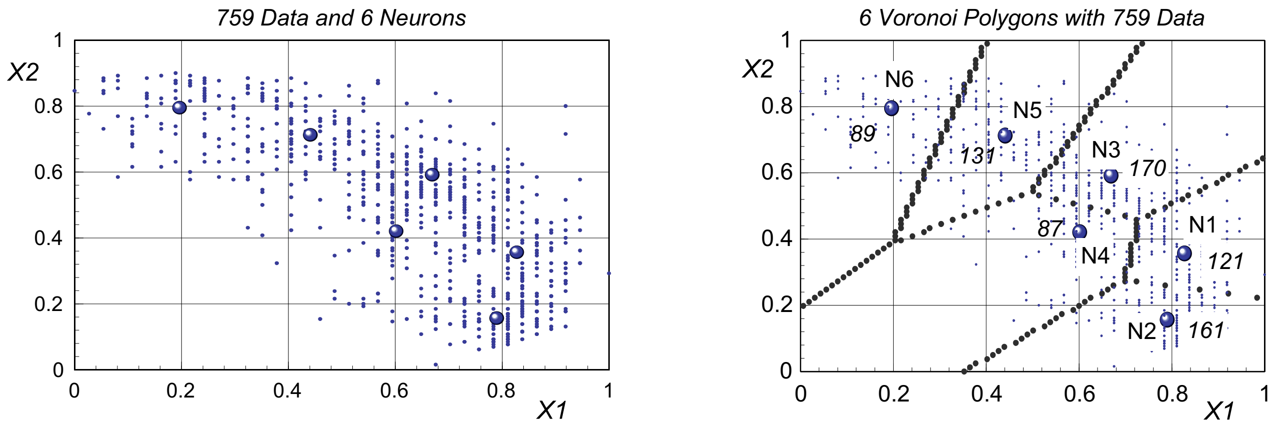


Figure 7. Information compression of *759* data with *6* Neurons and respective "*Voronoi polygons"*.

3.4 Computing the Neuron Widths W

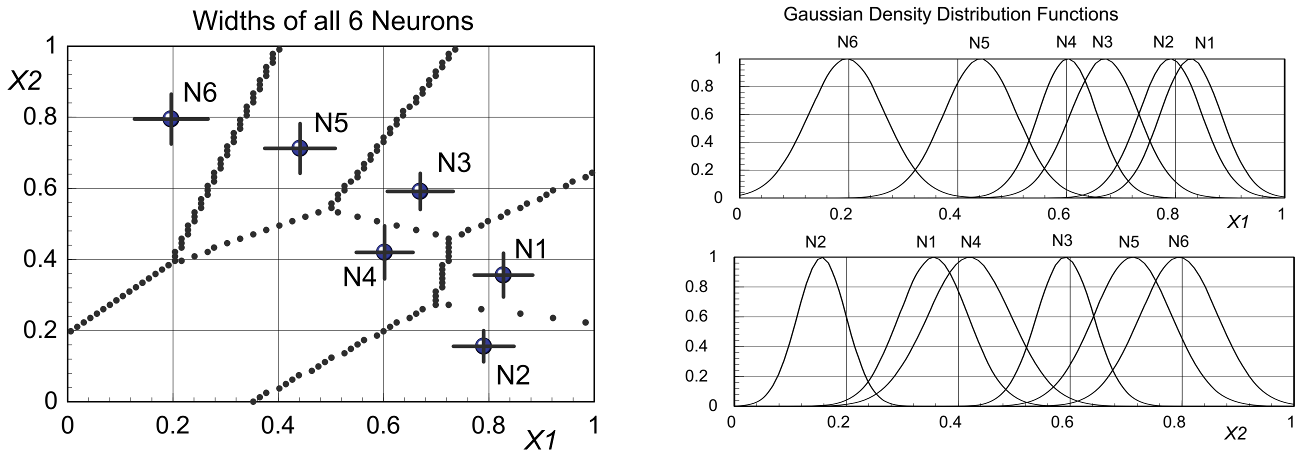
The size of the area covered by each "*Voronoi polygon*" does not reflect directly the data distribution in this area, since a part of it could be empty. This is clearly seen from the above Fig. 7. However, the neuron widths $\mathbf{W} = \{\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_N\}$ should provide a kind of evaluation of the data density distribution, which is important for the further step: information recovery (IR). Therefore, a plausible measure of such data distribution could be the average distance between the neuron center and all the data points in its "*Voronoi polygon*".

Computation of all the vectors for neuron widths $\mathbf{w}_n = [w_{n1}, w_{n2}, ..., w_{nK}],\ n=1,2,...,N$ is done by the following equation:

$$w_{nj} = \frac{1}{m_n} \sum_{s=1}^{m_n} |x_{sj} - c_{nj}|,\ j=1,2,...,K;\ n=1,2,...,N. \tag{10}$$

The following Fig. 8*a* is a scaled illustration of the widths of all 6 neurons, named as *N1, N2, ..., N6* from the example, shown in the previous Fig. 7. The bold horizontal and vertical lines around each neuron show their respective widths: $\{w_{n1}, w_{n2}\}$, $n = 1,2,...,6$ in the 2-dimensional parameter space $\{X1, X2\}$.

The computed neuron widths can be further regarded as widths of respective Gaussian density distribution functions, as shown in Fig. 8*b*. These functions are useful for approximate recovery of the density distribution of all *M* data over the whole range of the parameters, as it is shown in the following sections.



(a) (b)

Figure 8. Illustration of the neuron widths and the respective Gaussian density distribution functions.

4 METHODS FOR INFORMATION RECOVERY AND ANALYSIS OF DETERIORATION

The compressed information, consisting of the neuron centers $\mathbf{C}$, weights $\mathbf{G}$ and widths $\mathbf{W}$, as explained in the previous section are further transmitted to the Maintenance Center MC for information recovery (IR). There could be different methods for such recovery with their own merits and demerits, measured by the amount of the lost (unrecoverable) information. In this paper we propose and explain two relatively simple IR methods which can be used in the process of analysing the deterioration trends in the construction machines.

4.1 The Moving Window Average (MWA) Method

This simple method is used for recovering an important part of the information in the form of relationship between a given (predefined) *pair* of parameters, for example *P1 = X* and *P2 = Y*, selected from the group of all *K* parameters. The relationship can be drawn as a curve from a preliminary computed number of M_w points in the two-dimensional space $\{X, Y\}$, $\mathbf{V}_j = \{V_{jX}, V_{jY}\}$, $j = 1,2,...,M_w$. The number of these points coincides with the total of the windows M_w with a predefined size (length) *WL*, which are moved gradually from the left to the right boundary of the parameter *X* with a predefined overlapping *WO*. A graphical explanation of this Moving Window Average (*MWA*) method is given in Fig. 9*a* while Fig. 9*b* depicts an example based on real experimental data, taken from a hydraulic excavator.

The following abbreviations and notations are assumed in Fig. 10: *LB* – left boundary of the range of all neurons; *RB* – right boundary of this range; *B1* – left boundary of the current *j*-th window; *B2* – right boundary of the current *j*-th window (***W4*** in this case); *WL = B1 – B1*: window length; *WO* – window overlapping.

The most plausible way to evaluate the coordinates of the point $\mathbf{V}_j$ in the *j*-th window $(j = 1,2,...,M_w)$ is to compute the "*center-of-gravity*" of all N_j neurons that are "captured" in this window, as follows:

$$V_X = \sum_{i=1}^{N_j} C_{iX}^j G_i \Big/ N_j ; \;\; V_Y = \sum_{i=1}^{N_j} C_{iY}^j G_i \Big/ N_j ; \quad j = 1,2,...,M_w , \tag{11}$$

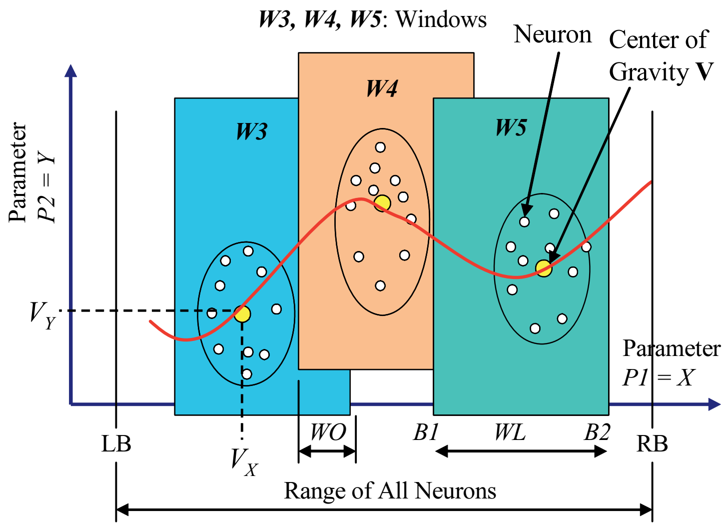
where C_{iX}^j and C_{iY}^j denote the center of the *i*-th neuron $(i = 1,2,...,N_j)$ and G_i is the relative weight of this neuron, computed by (9). Please note that as a result of this weighted computation of the coordinates (11), the point $\mathbf{V}_j$ does not fall in the geometrical center of the j-th window, but in its "*center-of-gravity*". Also, because of the overlapping between the windows, the following inequality holds: $\sum_{j=1}^{M_W} N_j > N$, where N is the total number of neurons.

The resolution of the computed relationship curve for the selected pair $\{X, Y\}$ of parameters depends on the window length *WL* as well as on the overlapping range *WO*. Generally, a bigger number of windows M_w and a moderate overlapping *WO* (less that *70 - 80*% between the neighbour windows) produce a better (smoother) curve without significant loss of information, compared with the performance of the Moving Window Average method on the original "raw data". The two close curves in Fig. 9*b* demonstrate such a good information recovery.

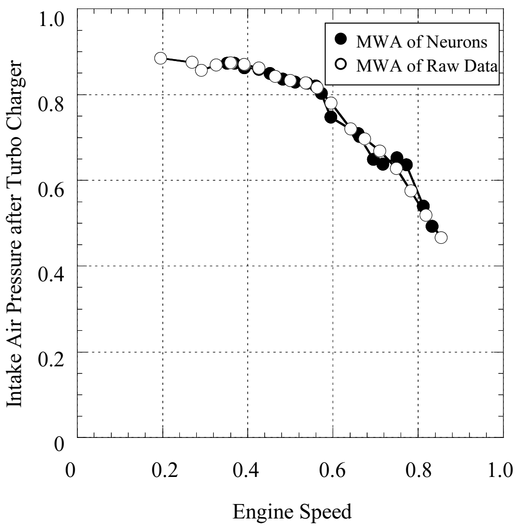
The MWA method could be used as an easy tool to discover differences between two or more operations of a machine, by visually comparing the relationships (curves) between certain pairs of parameters. Such example is provided in the last section of the paper.

4.2 The Method of Fuzzy Inference-based Relationship Analysis (FIRA)

The MWA method does not recover any information about the data density distribution, which could be quite important for discovering potential abnormalities in operation of the machine. The reason for this is that MWA method does not utilize the information about the data distribution, implicitly given by the neuron widths $\mathbf{W}$. Therefore here we propose another method, which is able to approximately evaluate (recover) the density distribution information. This method uses a fuzzy inference procedure to smoothly interpolate the density distribution along all parameters and will be referred to as FIRA method.



a) Graphical Illustration of the Moving Window Average Method;



b) MWA Performance, based on neurons and raw data.

Figure 9. Illustration of the MWA method and performance analysis of the method, based on "raw data" and on neurons.

The FIRA method tries to recover in the most plausible and smooth way the relative data density distribution along each parameter. Then a curve (in relative units) of the density distribution could be drawn for each parameter, for visualisation purposes, as shown in the next Fig. 10. It is worth to mention that the bold gray lines in Fig. 10., connecting the weights *G1, G2,…, G6* of the neurons for the two parameters *X1* and *X2* separately, do not represent in a true way the density distribution along the parameters, because they simply make a "piece-wise" linearization, which often causes "big jumps" between the far located (but neighbour) neurons, if there is a big difference in their weights (9).

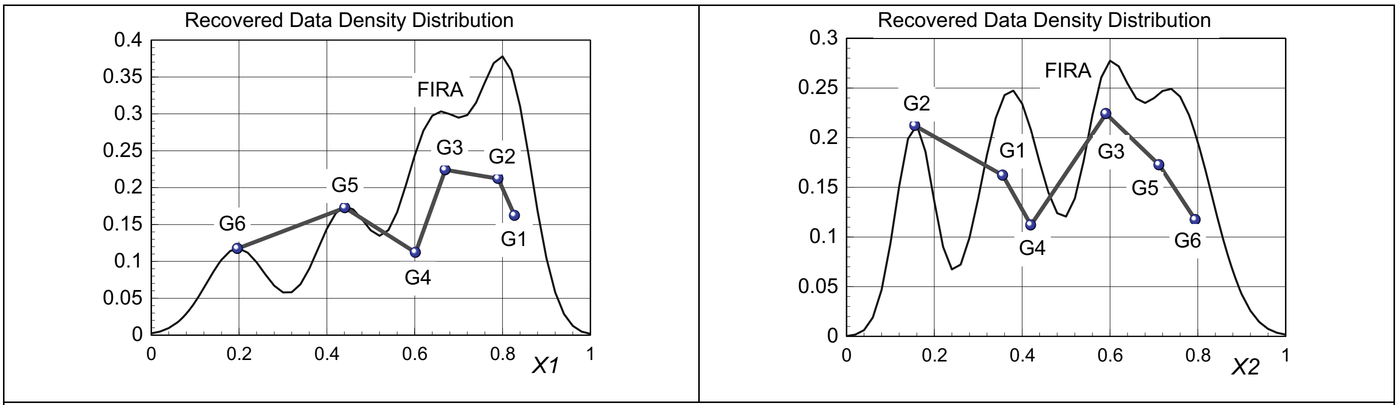


Figure 10. The recovered Data Density Distribution for the example in Figs. 7/8 by using FIRA method

Another way of using the FIRA method is to compute the so called "*Generalized Operating Point*" (GOP) which is one point $GOP = [x_o^1, x_o^2, ..., x_o^N]$ in the *K*-dimensional parameter space that shows the vicinity of the most frequent operation of the machine. Then a large number of GOPs in the parameter space can be computed from the compressed information for respective operations of one machine. The obtained "image" (shape, line) produced by the plotted GOPs (in each 2-dimensional plain) can be used for a simple visual detection of the existence of abnormal operations in this machine. Such operations will be represented by points (GOPs) being far from the general tendency (line) of the normal operations.

In the followings a brief explanation of the computational procedure for the FIRA method is given. As it will be seen, this method utilizes all parameters, obtained by the information compression procedure, namely neuron centers, weights and widths.

First of all, for each separate coordinate (parameter) r, $r=1,2,...,K$ we generate artificially (by scanning) *m* equidistant points, as follows: $x_j^r \in [P_{r\min}, P_{r\max}]$, $j=1,2,...,m.$ Then a "fuzzification" $0 \leq f_i(x_j^r) \leq 1$ of all points x_j^r for the current *r*-th parameter is made, as in the fuzzy sets theory and applications [4,5,7], by using *N* Gaussian membership functions (equal to the number of neurons) with their respective centers and widths:

$$f_i(x_j^r) = \exp\left[-\frac{(c_{ir} - x_i)^2}{2 w_{ir}^{\ 2}}\right], i=1,2,...,N; \ j=1,2,...,m. \tag{12}$$

The next step is to compute the *weighted density distribution* z_j^r in the *j*-th point x_j^r of the *r*-th parameter by using the fuzzified values from (12) and the weights of all *N* neurons , as follows:

$$z_j^r = \sum_{i=1}^{N} f_i(x_j^r) G_i, \ \ j=1,2,...,m. \tag{13}$$

Finally, each local coordinate of the GOP is computed as "weighted average", known also as "defuzzification" in the fuzzy sets theory in the following way:

$$x_o^r = \sum_{j=1}^{m} z_j^r x_j^r \Big/ \sum_{j=1}^{m} z_j^r, \ r=1,2,...,K, \tag{14}$$

The density distributions (the smooth curves) in Fig. 10 are computed by the above computation procedure. Some application examples for computing the GOP based on the FIRA method are given in the next section.

5 ANALYSIS OF DIFFERENT OPERATIONS OF A HYDRAULIC EXCAVATOR

In this section, the above presented methods for information recovery, namely MWA and FIRA methods, as well as the whole IC-IR technology for information compression and information recovery are illustrated and analysed by using real data from different operations of a hydraulic excavator, in order to detect possible deteriorations and abnormalities.

Six operations (considered as *normal*) of the excavator were available, taken during daily operation of the machine at different seasons of the year. In addition, data from seven other (*abnormal*) operations of the excavator, working on a special proving land, were also obtained. The abnormal operations were performed by emulating different types of faults in the diesel turbo-engine, by respective human intervention (special changes in the tuning of the engine parts and parameters). Among them are the following (most distinct) four faults: *F1* - Fuel Spray Nozzle of one cylinder is deactivated; *F2* – Deterioration in the Turbo-charger efficiency; *F3* – Air Filter Obstruction (*Big*) and *F4* - Air Filter Obstruction (*Small*). In order to achieve a fair comparison of the different operations, similar number of data points has been collected from each operation, in the range of *2000 - 2500* data with information compression made by using $N = 80$ neurons.

The following Fig. 11. depicts the case of the abnormal operation (fault *F1*). As seen from the comparison, based on the MWA method and shown in Fig. 12., this faulty operation can be easily detected, as the relationships between the pairs of parameters are quite different from those obtained from the normal operation with data, shown in the above Fig. 5. and Fig.6.

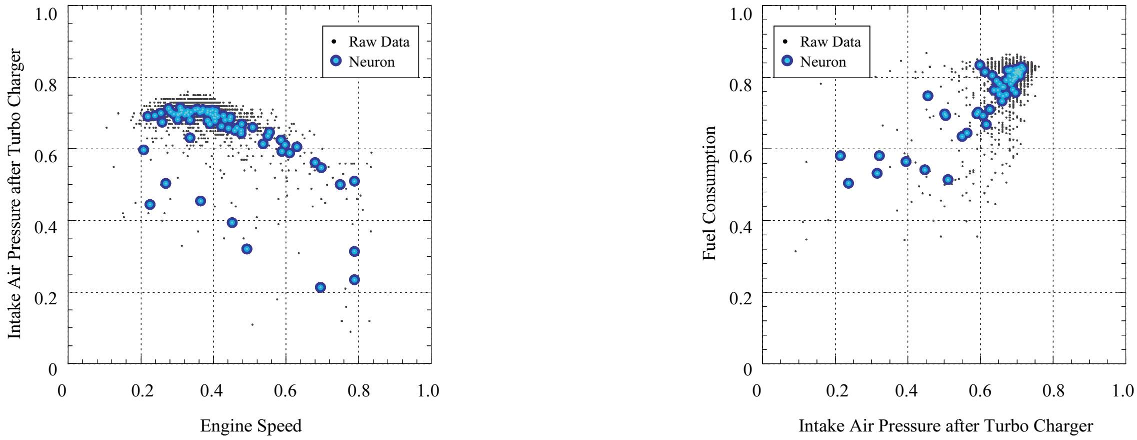


Figure 11. Plots of raw data and neurons in the case of abnormal operation (fault F1).

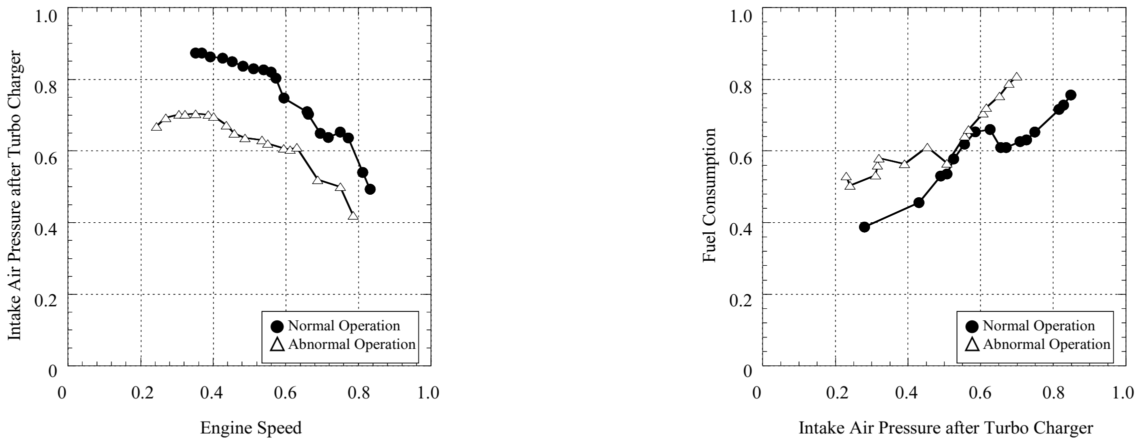


Figure 12. Detection of abnormal operation (fault *F1*) by using the MWA method

Finally, the FIRA method was used for all *6* normal and *7* abnormal operation data in order to create the plots of the *13* generalized operating points (GOPs), shown in the next Fig. 13. The abnormal operations are shown as diamond curve symbols, while the normal operations and denoted by bold circles.

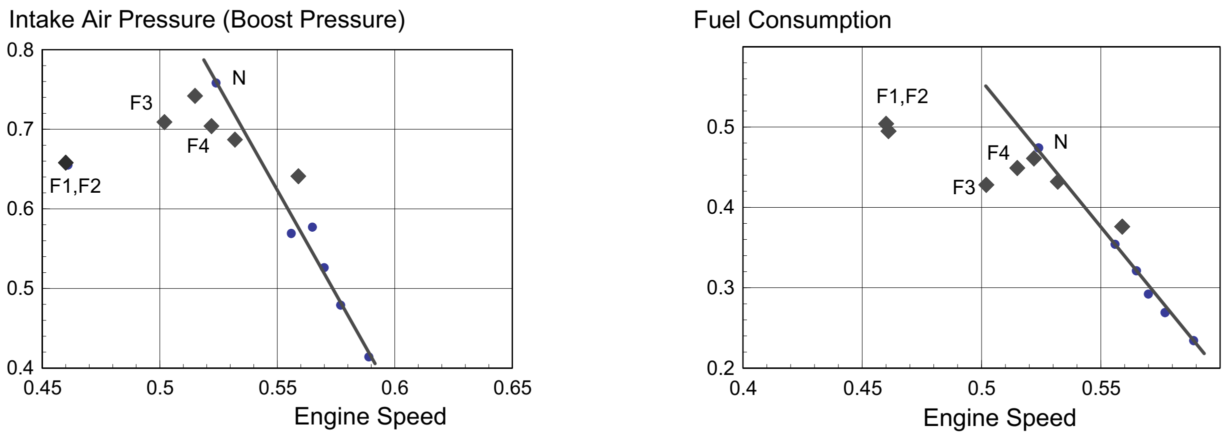


Figure 13. Detection of abnormal operations by using FIRA method, based on *6* normal and *7* faulty operations.

It is easy to notice from Fig. 13. that all *six* normal operations (including the operation denoted by *N*, which corresponds to the data from Fig. 5. and Fig.6.) are characterized by a straight line in the *2*-dimensional parameter space, which means that the working regime (performance) of the machine is not changed. The big distances between some of the operating points GOPs are explained with the *different load,* under which the machine was normally operating on that day.

It is also seen from te above Fig. 13. that most of the faulty operations (including the "strong" faults *F1,F2,F3* and *F4*) stay apart from this "normal operation line". The remaining four faults are considered as "mild" faults, i.e. they are very similar to the normal operation of the machine, because their GOPs are close to the "normal operation line".

6 CONCLUSIONS

The IC-IR technology and system for information compression and information recovery, proposed and analysed in this paper is an efficient tool for processing and dealing with large amount of data taken from the daily operation of the construction machines for the purpose of their performance evaluation, fault detection, diagnosis and maintenance. The real data used from different (normal and faulty) operations of a hydraulic excavator and their successful analysis prove the system applicability. Currently the whole system and the algorithms are under further refinements in order to improve the resolution level of the IC-IR process, as well as to make it more practical for early automatic detection of abnormal and faulty states in machines.

7 REFERENCES

[1] Gertler J. (1999) Fault Detection and Diagnosis in Engineering Systems. New York: Marcel Dekker.

[2] Quing Zhao (2004) Design of a Novel Knowledge-Based Fault Detection and Isolation Scheme, IEEE Trans. SMC -B , 34(2), 1089-1095.

[3] Maki Y. & Loparo KA. (1997) A Neural-Network Approach to Fault Detection and Diagnosis in Industrial Processes, IEEE Trans CST, 5, 529-541.

[4] Isermann R. (1998) On Fuzzy Logic Applications for Automatic Control, Supervision and Fault Diagnosis, IEEE Trans. SMC – B, 28, 221-235.

[5] Pedrycz W. (2005) Knowledge-Based Clustering. From Data to Information Granules. A John Wiley & Sons Inc., Publication, Printed in USA.

[6] Vachkov G., Kiyota Y & Komatsu K. (2004) Real-Time Classification Algorithm for Recognition of Machine Operating Modes by Use of Self-Organizing Maps, Turkish Journal of Electrical Engineering & Computer Science ELEKTRIK, 12(1), 27-42.

[7] Vachkov G., Kiyota Y & Komatsu K. (2005) Detection of Changes in System Performance by Use of RBF Network Models and Fuzzy Rules, Proc. of the Int. Conf. on Mechantronics and Automation ICMA 2005, Niagara Falls, Ontario, Canada, 608-615..

[8] Bishop CM. (2003) Neural Networks for Pattern Recognition. Oxford University Press, Published in the United States. .

[9] Kohonen T. (2001) Self-Organizing Maps, Third Edition, Springer Series in Information Sciences, Springer, Berlin.

[10] T. Martinetz T, Berkovich S & Schulten K. (1993) Neural-Gas Network for Vector Quantization and Its Application to Time-Series Prediction, IEEE Trans. Neural Networks, 4(4), 558-569.

A STATISTICAL ACTIVITY COST ANALYSIS OF THE RELATIONSHIP BETWEEN PHYSICAL AND FINANCIAL ASPECTS OF FIXED ASSETS[1]

A. Colin[†] M. Falta[†2] S. Su[†], L. Turner[‡] R. Willett[†] & R Wolff[‡]

School of Accountancy[†] & School of Economics and Finance[‡] Queensland University of Technology, Brisbane QLD 4001, Australia

Abstract: In this paper, Statistical Activity Cost Analysis (SACA) is used to identify the interaction of mutually dependent physical and financial aspects of a fixed asset-like system configuration. The novelty of the approach is, having established a rational description of the uncertainty inherent in both domains, the analysis of their interaction. Little research to date has investigated the duality of engineering and accounting aspects, in a statistical setting. Our approach is conceptual rather than empirical. We use an illustrative 4-component model, a) to explain the concept of SACA by means of a software demonstration tool, b) to relate financial issues of cost to engineering asset capacity to perform specified tasks, and c) to demonstrate how to produce quantified measures of return and risk, both of which are relevant in areas of life-cycle analysis, budgeting and planning decision-making.

Key Words: Cost Accounting; Maintenance; Reliability; Risk and Risk Attribution; Statistical Activity Cost Analysis (SACA); Uncertainty;

1 INTRODUCTION

The motivation for the research upon which this paper is based is to scientifically model the linkages between economic and physical processes. This is necessary to enable us to relate engineering decisions regarding the management of physical assets to financial consequences in terms of return and risk. It is also necessary to understand how financial decisions effect physical asset capacity to perform specific tasks.

Little research to date has investigated the effects of cost uncertainty and its variation in conjunction with the physical operation of fixed assets. The ability to identify sources of uncertainty which require more attention than others and the quantification of the risk in areas of planning, budgeting or life-cycle analyses facilitate better decision-making. In this paper we advocate that the cost variable in engineering models needs to be a random variable to achieve this goal. Observed costs in invoices or transaction's data give us realisations of this random variable.

In accounting, a rational framework which provides such an interpretation of costs is Statistical Activity Cost Analysis (SACA) due to Willett (1987, 1988, 1991) and extended by Gibbins & Willett (1997) and Falta & Wolff (2004). To date, applications of this framework to aspects of reliability theory, maintenance or life-cycle analysis have been limited (Falta, 2005).

From the perspective of engineering, physical aspects of assets described by concepts of reliability and maintenance are well-understood (e.g., Kumar, Crocker, Knezevic & El-Haram, 2000). However, the literature that considers the uncertainty of cost and physical asset properties *jointly* is more sparse. Most literature in engineering and operations research assumes costs are a directly observable, and in this sense, deterministic attribute of maintenance, production or usage processes. Compare, for example, an optimisation procedure for costs and safety considerations on component systems (Vaurio, 1995), statistical life-cycle cost analyses (e.g., Jiang, Zhang & Ji, 2003; Monga & Zuo, 2001), literature that targets specific assets such as water mains (e.g. Engelhardt, Skipworth, Savic, Saul & Walters, 2000), costs, efficiency and environmental implications of gas pipelines (e.g., Bergerson & Lave, 2005) or railway maintenance (e.g., Lamson, Hastings & Willis, 1983). In the operations research literature, costs of physical aspects of assets have been investigated, for example, in relation to software reliability (Pham, 2003), machine replacement decisions (e.g., Dogramaci & Fraiman, 2004), policy-making for replacement or standby decisions (e.g., Hsieh, 2005).

[1]This work is sponsored by the Australian DSTO (Defence Science and Technology Organisation) through the Australian CRC for Integrated Engineering Asset Management as Projects MD101 and MD105.

[2]Corresponding author: m.falta@qut.edu.au

Common to the above literature is the lack of a simple generic model that relates physical characteristics to the uncertainty and variation of corresponding costs that result from dependencies in the physical and financial dimensions of process. Using the principal *representation theorem* of SACA, which explicitly superimposes the financial structure of business processes on the input-output relationships of activities in the physical structure, costs can be modelled as *random sums of random variables*. This lays the basis for a comprehensive analysis of the financial risk involved in engineering processes.

In Section 2 of this paper we employ some fundamental concepts of the SACA framework, based on ideas given in Vaurio (1995), in a system of four components that are in a series constellation. Our focus in this section lies with the demonstration of these theoretical concepts by means of the *SACA Concept Demonstrator*. We conclude in Section 3 with some brief final remarks.

2 THE SACA CONCEPT DEMONSTRATOR

We have developed an illustrative software tool called the SACA Concept Demonstrator in order to better visualise the capabilities of the SACA approach to relate costs to engineering asset capability[3]. Let, for example, an asset have four components functioning in series with their failure characteristics described by exponential time-to-failure distributions with parameters λ_i, i=1,..,4, for each component i. After each maintenance operation on a component with expected cost C_i let the system be as-good-as-new and the MTTF (Mean Time To Failure) be $1/\lambda_i$.

For this simple hierarchical system, the SACA Concept Demonstrator is applied to two scenarios in the first of which we minimise the maintenance costs given an *a priori* minimum reliability level. In the second scenario we maximise the reliability of the system, given a budget.

2.1 Scenario 1: Cost Minimisation with Set System Reliability R_{min}

Assume that maintenance operations on a component i may be carried out at any time and at any frequency f_i. Then the total cost of maintaining the system over a single interval s is

$$C_s = \sum_i \frac{C_i}{f_i} \tag{2.1}$$

We achieve minimum costs C_s, considering for the system reliability $R_s \geq R_{min}$, by optimising f_i in expression (2.1). Using the method of Lagrange multipliers and after some calculus (cf. Appendix A) yields an explicit expression for f_i in terms of λ_i, C_i and R_{min},

$$f_i = -\left(\frac{C_i}{\lambda_i}\right)^{1/2} \cdot \frac{\ln\{R_{\min}\}}{\sum_i (\lambda_i C_i)^{1/2}} \tag{2.2}$$

In Table 1, we have displayed the results from a simulation study with R_{min}=0.97. The overall minimum cost is C_s=\$2187.28 (sum of column (6)). Using Microsoft Excel's Solver function for the same optimisation problem yields for C_s=\$3606, 39% higher than the cost calculated by the present method.

Table 1 Simulation Outputs.

(1) Component	(2) C_i	(3) λ_i	(4) f_i	(5) # Maintenance Operations	(6) Total Costs	(7) Reliability
1	\$10	0.1	0.03732	26.7974	\$267.97	0.9963
2	\$20	0.2	0.03732	26.7974	\$535.95	0.9926
3	\$10	0.4	0.01876	53.5948	\$535.95	0.9926
4	\$20	0.5	0.02360	42.3704	847.41	0.9883

Results from simulation for a four component system in series for given maintenance costs C_i, parameters λ_i and minimum system reliability R_{min}=0.95. Results from using expression (2.2) yield maintenance frequencies f_i, the number of maintenance operations per interval and respective total maintenance costs per component, and the (minimum) component reliabilities.

In the SACA Concept Demonstrator, the expected maintenance cost and its variance, assuming the Normal as the underlying distribution, for each component can be altered and so can the MTTF parameters, assuming exponential reliability

[3]Asset capability is a somewhat general expression and refers to an asset's ability to perform certain tasks. We use the reliability function to describe engineering asset capability, the physical condition of an asset.

functions. In practical applications, both cost parameters can be either estimated from observed data in form of invoices or transaction statements or fitted by theoretical distribution functions. A starting value for the MTTF parameters is usually provided by the manufacturer which, if feasible, can be later adjusted by the operator.

Table 2 SACA Concept Demonstrator Output.

Expected Maintenance Costs [$]	MTTFs [*month*]	Optimal Maintenance Frequency f_i at R_{min}=0.95
20/20/20/20	8/13/6/8	10/8/11/10
20/20/20/20	8/14/6/8	10/7/11/10
20/20/20/20	8/15/6/8	9/7/11/9
26/18/37/20	8/9/6/8	10/11/10/11
26/18/41/20	8/9/6/8	10/11/9/11
26/18/45/20	8/9/6/8	10/12/9/12

Results from variations of parameter settings of expected maintenance costs and MTTFs of single components. The results are displayed in the following form: component 1/component 2/component 3/component 4.

For each setting of expected maintenance costs and MTTFs of the system, the SACA Concept Demonstrator calculates the optimal number of maintenance operations per interval (rounded-up numbers) at minimum costs with a reliability threshold of 0.95. The calculations are performed according to expression (2.2) and the results in the SACA Concept Demonstrator are displayed according to single lines in Table 2. Note that due to the character of expression (2.2), the MTTF of one component is reciprocally correlated with the f_is of every other component but positively with its own f_i. Thus, increasing, for example, λ_2 and maintaining constant expected costs for all components might change f_1, the number of maintenance operations for component 1. Compare Table 2, rows 2 to 4 for this example. Similarly, because the expected maintenance cost C_i of one component is negatively correlated with the expected costs of all other components for constant MTTFs, increasing C_3, for example, may increase f_i, i=1,2,4, with respect to the other components while decreasing f_3, the number of maintenance operations performed on the component itself. Compare Table 2, rows 5 to 7 for this example.

2.2 Scenario 2: Reliability Maximisation with a Set Budget

In the second scenario, we use the same four component system as above, however, with a given fixed budget B. It is now of interest as to how one should allocate the maintenance frequencies in order to derive the highest possible reliability.

Expression (2.1) becomes

$$C_s = \sum_i \frac{C_i}{f_i} - B, \tag{2.3}$$

and we derive the optimal maintenance frequencies according to

$$f_i = -\left(\frac{C_i}{\lambda_i}\right)^{1/2} \cdot \frac{\sum_i (\lambda_i C_i)^{1/2}}{B}. \tag{2.4}$$

Details of the derivation of expression (2.4) are given in Appendix B.

In Table 3, some results from simulation are displayed for a budget of $1500. The maintenance frequencies that give the budget at maximum reliability are derived in the first column of the table. Note that the product of all component reliabilities has decreased to 0.9565.

2.3 The SACA Concept Demonstrator Displays

2.3.1 Reliability vs. Cost

From the above calculations, it is straightforward to plot a curve for the total cost C_s at any level of reliability. This curve is the dark, continuous line displayed in Figure 1. The impact of an increase in expected maintenance costs for single components increases with increased reliability, as would be expected (cf. left and right panel in Figure 1). In the left panel, the expected maintenance cost for components 1 to 4 are $15, $20, $12 and $39, and in the right panel, $26, $20, $19 and $39, respectively.

Table 3 Simulation Outputs.

(1) Optimal Maintenance Frequency f_i	(2) # Maintenance Operations per Interval	(3) Costs	(4) Reliability
0.054415	18.38	$183.77	0.994573
0.054415	18.38	$367.54	0.989176
0.027208	36.75	$367.54	0.989176
0.034415	29.06	$581.14	0.98294

Maintenance frequencies that yield maximum system reliability, given a fixed budget on maintenance.

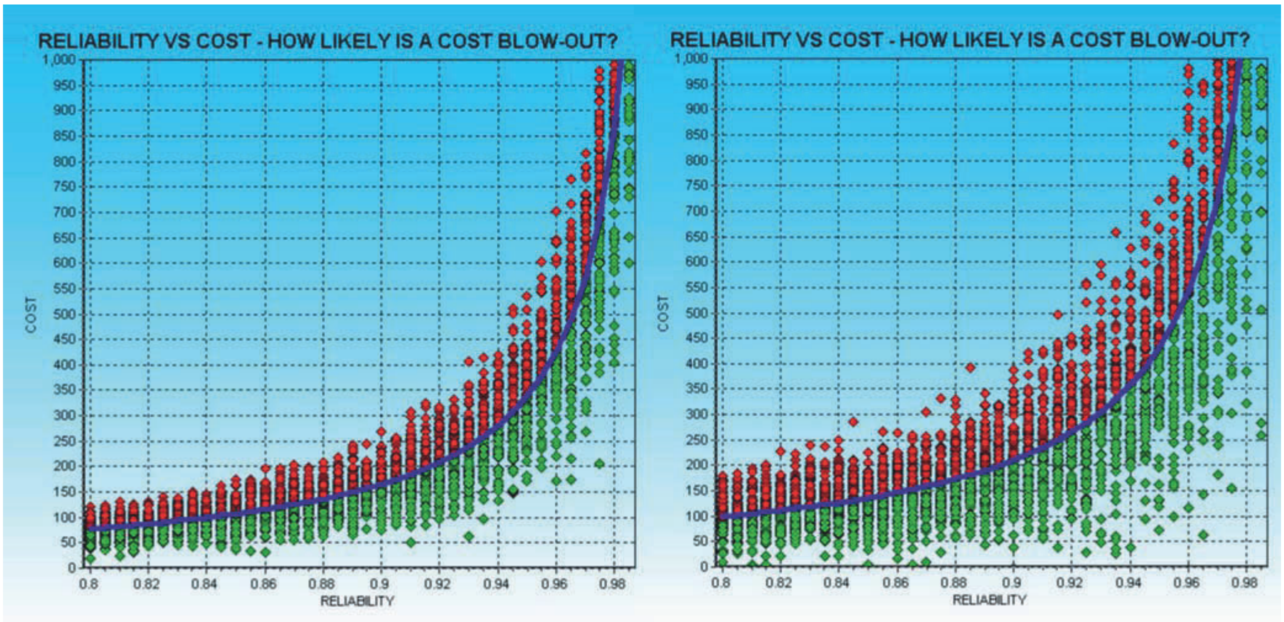


Figure 1: Printed screens from the SACA Concept Demonstrator

Curves for reliability level versus total maintenance costs (dark, continuous line). The scatter around the curve results from the variance of maintenance costs and can be used as an indicator of the risk for cost overruns (above the line) or cost savings (below the line).

The scatter around the reliability versus cost curve derives from the variation in maintenance costs, a parameter that can be varied. In Figure 1, the cost over and underruns are displayed by single dots at each reliability level for a given set of variances of maintenance costs. Notice that the increase in spread in the right plot of the figure, which is due to an increase from $6 to $12 of the variance of maintenance costs of component 1, is rather sensitive. All other parameters remained constant.

In practice, it is usual that the quality of estimates for maintenance and repair costs varies, for some are robust and others more vague. Therefore, visualisation and quantification of the likely cost overruns will be of utmost importance for the decision-maker.

2.3.2 Risk Attribution

The functionality for risk attribution is displayed on the top panel of Figure 2. The pie charts display the percentage contribution of single components to achieve the stated budget at a confidence level of 95% and given reliability.

Top panel: the calculated budget is based on the optimal number of maintenance operations for a given set of simulation parameters. The pie charts display the source of cost uncertainty from each component in order to achieve that budget at a 95% confidence level for a given level of reliability. An increase in the cost variability of component 1 from $6 to $12 increases the necessary budget to guarantee the given reliability at the 0.95 level.

Bottom panel: quarterly upper 95% limits corresponding to the pie charts in the same column above. Numbers that add up to the total number of optimal maintenance operations for each component have been assigned to each quarter according to the simple rule described in the text.

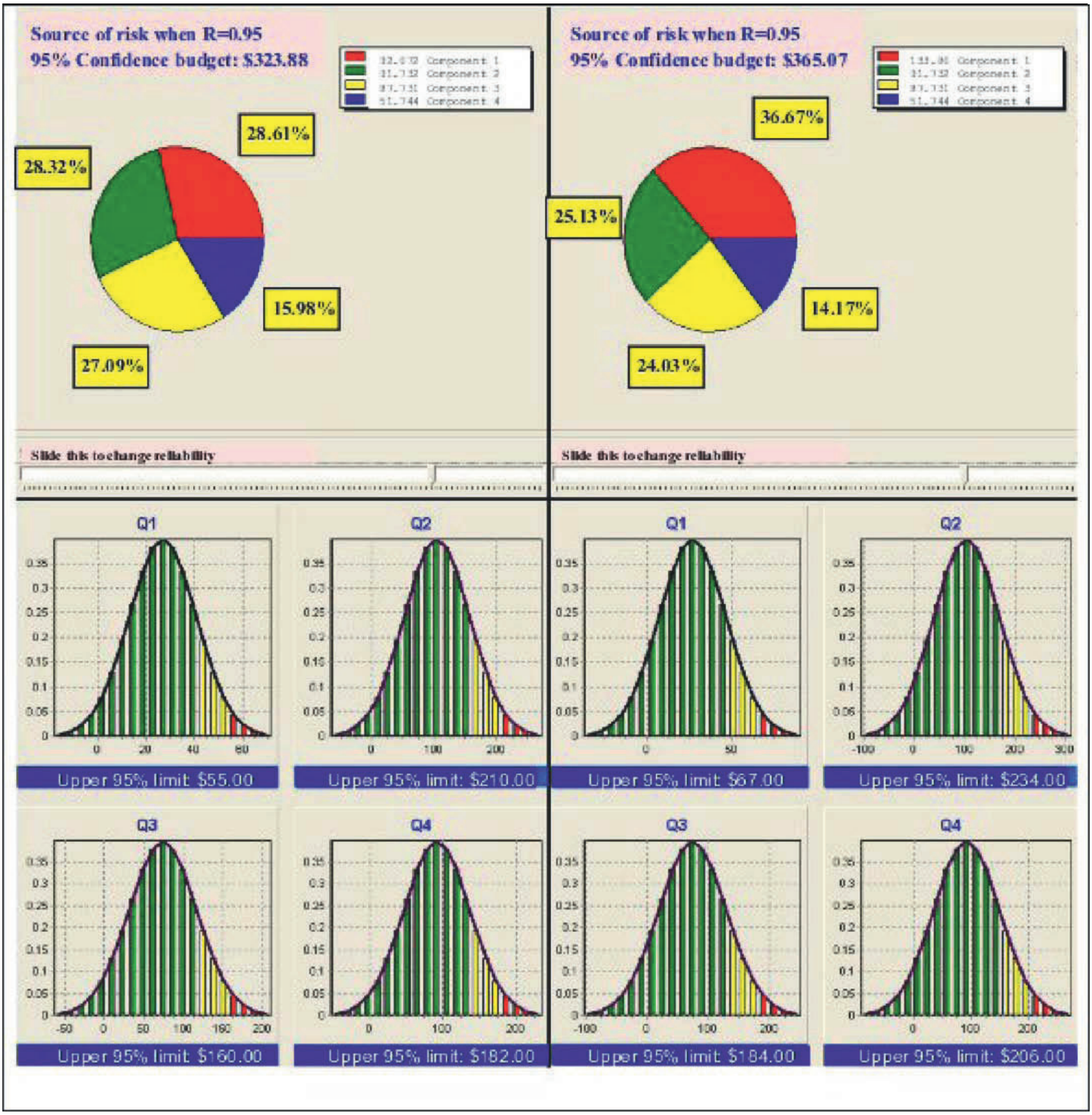


Figure 2 Printed screens from the SACA Concept Demonstrator

Thus, the decision-maker is able to assess the source of risk and quantify it. The parameter settings for the left pie chart are, for components 1 to 4, expected maintenance costs of \$15, \$20, \$12 and \$30; variances of \$6, \$15, \$8, and \$10; MTTF parameters of 8, 21, 12, and 44 months; and calculated numbers of maintenance operations of 7, 4, 7 and 2, respectively. The difference in the pie charts results from an increase (\$6 to \$12) of the variance of maintenance costs of component 1. The increase in variability of this component has not changed the optimal number of maintenance operations. However, the impact on numbers in the pie chart is significant: to guarantee, at a 95% level of confidence, that reliability remains above 0.95, a budget of \$365.07 has to be invested as opposed to \$323.88, as displayed in the left top panel. For this 12.72% increase in the budget necessary to maintain the reliability of the system, all the additional resources are consumed by component 1 (\$290.09 instead of \$200.83) and is accompanied by an increase in uncertainty from 28.61% to 36.67%. The costs of components other than component 1 remain constant with their percentage uncertainty decreasing.

2.3.3 Quarterly Budgets

In the bottom panel of Figure 2, the quarterly upper 95% confidence limits are displayed (dark bars on the right of the light bars) for identical parameter settings as shown in the pie charts. The four graphs on the left bottom panels correspond to the pie chart in the top left panel. The same pattern applies to the right side of the figure. Note the different scaling of the abscissae. The varying numbers for the quarterly risk thresholds result from assigning the given numbers of optimal maintenance operations N_i, i=1,..,4, as follows: the numbers N_i are distributed across the interval (year) at equidistant durations.

This display holds two different pieces of information for the decision-maker. Firstly, for every single period the risk of cost overruns is displayed. In organisations where budget allocations across divisions or projects are a matter of reality, the funds receiving party is able to demonstrate, on a rational basis, the adequateness of any amount allocated. Secondly, the series of amounts at the confidence limit sheds light into the time series of maintenance costs. Actual maintenance costs correct the estimates for passed time intervals and update the estimates for future time intervals. The latter point, including to account for other time intervals than quarterly periods, leads to the world of time series and financial option analysis, which we leave for future research.

2.4 Comments on the SACA Concept Demonstrator

We have demonstrated how the SACA framework can be applied to the financial assessment of a physical system and, in particular, how the cost variations can be expressed by statistical distributions. We have furthermore demonstrated how the results can be used by the decision-maker for maintenance scheduling and budget planning.

The underlying model can be extended in various directions. Below we have listed and elaborated on some of these.

1. The assumption that maintenance may be carried out at any time and at any frequency is a limitation when considering practical applications. Usually there are scheduled maintenance windows or maintenance opportunities at discrete time intervals of set durations. It is a matter of choice whether these constraints should drive the simulation, or whether the results of such simulations should be used to modify the time and frequency of operations.
2. Greater detail in the analysis of the cost-reliability relationship can be achieved by allowing for the general *k*-out-of-*n* parallel and series systems. Our choice to consider exponential time-to-failure distributions is based on the following. The illustrative 4-component model can be viewed as a reduced representation of a more complex reliability block diagram. It is usually the case that the larger the number of components and actors in an entity the larger the number of failure modes, and as a consequence the more random and unpredictable the times of failure. The exponential time-to-failure distribution models this type of situation because it implies a Poisson distribution for the number of failures during equal time intervals. For empirical evidence we refer to, for example, Davis (1952) who concluded that "the exponential theory of failure appears to describe most of the systems examined [in the article]". Any other time-to-failure distribution such as the Weibull distribution, a popular choice among engineering practitioners, may be chosen if it is appropriate to do so. A further remark on this topic is given in Point 5.
3. The lower bound used on reliability levels in the model assumes that all components are functioning at their lowest reliability, even if a maintenance operation has just been performed. A time-dependent reliability function can describe a more refined physical description of the system.
4. The resulting cost *vs.* reliability curve displayed in Figure 1 is continuous and represents the expected costs. This curve can be seen as the effective frontier, a concept used in portfolio analyses, as arbitrarily higher amounts of dollar may be spent on the maintenance of a component without improving the total system reliability, at any level. This optimisation problem is being currently investigated employing a genetic algorithm and is in preparation to accompany the current paper.
5. For other distributions that describe the component reliability, particularly those with more complex forms such as those with a bath-tub shape, numerical simulations might be the first choice for parameter estimates.

3 FINAL REMARKS

We have used a simple four component system in order to demonstrate new concepts in the field of accountancy. Invoices of maintenance operations are modelled using statistical distributions, which describe the uncertainty of varying costs incurred by businesses. For some of the components expected costs can be determined quite accurately from prior experience, others have to be estimated from observed data. The variation resulting from the former suppositions and latter estimates contributes to the overall risk of budget overruns. With the SACA Concept Demonstrator, a simple module with a graphical user interface, implements the methods described in the paper. In doing this and in further developments, we intend to bridge the fields of accountancy (finance), reliability theory (physical) and maintenance (policy) to improve fixed asset management.

Of the limitations in the model, of most immediate concern to us are the need to extend the physical model to *k*-out-of-*n* parallel redundant systems and to use observed data, which necessitates the use of reliability distributions that describe actual component behaviour.

This paper illustrates the applicability of the SACA framework to an assessment of the return and risk linkages between decisions regarding physical asset properties and financial outcomes, *and* between financial decisions and physical outcomes. The framework can be applied to any asset and typically requires information to be extracted from both accounting and physical asset management systems. At the present time this approach to modelling is being applied to assets of the Royal Australian Navy as part of the Logistics Cost of Ownership initiative of the Australian DMO (Defence Materiel Organisation). It is also being applied in the Australian food processing and utilities industry.

Appendix A

For a series system, consider constraint

$$R_{\min} \le R_s = \prod_i R_i = \prod_i \exp(-\lambda_i T_i). \tag{A1}$$

Since $T_i \ge 1/f_i$ a lower bound on each component's reliability is given by $R_i \ge \exp(-\lambda_i f_i)$ and expression (A1) is violated if

$$R_{\min} > \prod_i \exp(-\lambda_i f_i) \tag{A2}$$

implying that

$$\ln\{R_{\min}\} > -\sum_i (\lambda_i f_i). \tag{A3}$$

Note that expression (A3) is unnecessarily strong since it assumes that the lowest component reliabilities will always hold.

Expressions (2.1) and (A3) together form a constrained optimisation system which we solve using Lagrangian multipliers. Thus, combining constraint and target expressions respectively as

$$f_c(f_i) = \sum_i \lambda_i f_i + \ln\{R_{\min}\} = 0 \tag{A4}$$

$$f_t(f_i) = \sum_i \frac{C_i}{f_i}, \tag{A5}$$

so that

$$\frac{\partial f_c}{\partial f_i} = \lambda_i \quad \text{and} \quad \frac{\partial f_t}{\partial f_i} = -\frac{C_i}{f_i^2}. \tag{A6}$$

The solution to

$$\frac{\partial f_t}{\partial f_i} + \hat{L}\frac{\partial f_c}{\partial f_i} = 0 \tag{A7}$$

is an extremum of the function f_t subject to the constraint $f_c = 0$, where $\hat{L}$ is a Lagrange multiplier. We can therefore reformulate this problem by writing

$$\frac{C_i}{f_i^2} = \hat{L}\lambda_i. \tag{A8}$$

The system of equations in *n+1* unknowns represented by expressions (A4) and (A8) is solved by substituting the implied expression for f_i in (A8) back into (A4) yielding

$$\sum_i f_i \left(\frac{\lambda_i C_i}{\hat{L}}\right)^{1/2} + \ln\{R_{\min}\} = 0 \tag{A9}$$

which gives

$$\hat{L} = \left(\frac{\sum_i \lambda_i C_i}{-\ln\{R_{\min}\}}\right)^2. \tag{A10}$$

Substituting the last expression into (A8) yields expression (2.2) of Section 2.1.

Appendix B

Similar to expressions (A4) and (A5) we have functional forms for constraint and target, respectively,

$$f_c(f_i) = \sum_i \frac{C_i}{f_i} - B = 0 \quad \text{and} \quad f_t(f_i) = \sum_i \lambda_i C_i. \tag{B1}$$

Thus,
$$\frac{\partial f_t}{\partial f_i} = \lambda_i \quad \text{and} \quad \frac{\partial f_c}{\partial f_i} = -\frac{C_i}{f_i^2}. \tag{B2}$$

and we solve the expression

$$\lambda_i = \hat{L}\frac{C_i}{f_i^2} \tag{B3}$$

which, substituting for f_i back into the first equation of expression (B1) yields

$$B = \sum_i \left(\frac{\lambda_i C_i}{\hat{L}}\right)^{1/2}. \tag{B4}$$

Substituting the resulting value of $\hat{L}$ into expression (B3) yields expression 2.4 of Section 2.2.

4 BIBLIOGRAPHY

[1] Bergerson, J. A., & Lave, L. B.; **2005**; "Should We Transport Coal, Gas, or Electricity: Cost, Efficiency, and Environmental Implications"; *Environmental Science & Technology*; 39, 16, 5905

[2] Davis, D. J.; **1952**; "An Analysis of Some Failure Data"; *Journal of the American Statistical Association*; 47, 258, 113-150

[3] Dogramaci, A., & Fraiman, N. M.; **2004**; "Replacement Decisions with Maintenance Under Uncertainty: an Imbedded Optimal Control Model"; *Operations Research*; 52, 5, 785

[4] Engelhardt, M. O., Skipworth, P. J., Savic, D. A., Saul, A. J., & Walters, G. A.; **2000**; "Rehabilitation Strategies for Water Distribution Networks: a Literature Review with a UK Perspective"; *Urban Water*; 2, 153-170

[5] Falta, M.; **2005**; "Statistical and Computational Methods to Assess Uncertainty and Risk in Accounting"; http://adt.caul.edu.au

[6] Falta, M., & Wolff, R.; **2004**; "Recent Developments of Statistical Approaches in Aspects of Accounting: a Review"; *International Statistical Review*; 72, 3, 377-396

[7] Gibbins, M., & Willett, R. J.; **1997**; "New Light on Accrual, Aggregation and Allocation, Using an Axiomatic Analysis of Accounting Numbers' Fundamental and Statistical Character"; *ABACUS*; 33, 2; 137-167

[8] Hsieh, C. C.; **2005**; "Replacement and Standby Redundancy Policies in a Deteriorating System with Aging and Random Shocks"; *Computers & Operations Research*; 32, 9, 2297

[9] Jiang, R., Zhang, W. J., & Ji, P.; **2003**; "Required Characteristics of Statistical Distribution Models for Life Cycle Cost Estimation"; *Int. J. Production Economics*; 83, 185-194

[10] Kumar, U. D., Crocker, J., Knezevic, J., & El-Haram, M.; **2000**; "Reliability, Maintenance and Logistic Support: A Life Cycle Approach"; Kluwer Academic Publishers

[11] Lamson, S. T., Hastings, N. A. J., & Willis, R. J.; **1983**; "Minimum Cost Maintenance in Heavy Haul Rail Track"; *The Journal of the Operational Research Society*; 34, 3, 211

[12] Monga, A., & Zuo, M. J.; **2001**; "Optimal Design of Series-Parallel Systems Considering Maintenance and Salvage Value"; *Computers & Industrial Engineering*; 40, 323-337

[13] Pham, H.; **2003**; "Software Reliability and Cost Models: Perspectives, Comparisons, and Practice"; *European Journal of Operational Research*; 149, 3, 475

[14] Vaurio, J. K.; **1995**; "Optimization of Test and Maintenance Intervals Based on Risk and Cost"; *Reliability Engineering and System Safety*; 49, 23

[15] Willett, R. J.; **1987**; "An Axiomatic Theory of Accounting Measurement"; *Accounting and Business Research*; 155-171

[16] Willett, R. J.; **1988**; "An Axiomatic Theory of Accounting Measurement - Part II"; *Accounting and Business Research*; 19, 73; 79-91

[17] Willett, R. J.; **1991**; "Transaction Theory, Stochastic Processes and Derived Accounting Measurement"; *ABACUS*; 27, No. 2; 117-134

APPLICATION OF THE GENERALISED LIKELIHOOD RATIO ALGORITHM TO THE DETECTION OF A BEARING FAULT IN A HELICOPTER TRANSMISSION

F. Antonio Galati [a], B. David Forrester [a] and Subhrakanti Dey [b]

[a] *Air Vehicles Division, Defence Science and Technology Organisation, 506 Lorimer Street Fishermans Bend, VIC 3207, Australia.*

[b] *Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, VIC 3010, Australia.*

Abstract: A Bell 206B main rotor gearbox was run at high load under test conditions in the Helicopter Transmission Test Facility operated by the Defence Science and Technology Organisation (DSTO) of Australia. The test succeeded in initiating and propagating pitting damage in one of the planet gear support bearings. Vibration acceleration signals were recorded periodically for the duration of the test. The time domain vibration signals were converted to angular domain to minimise the effects of speed variations. Auto-Regressive Moving-Average (ARMA) models were fitted to the vibration data and a change detection problem was formulated in terms of the Generalised Likelihood Ratio (GLR) algorithm. Two different forms of the GLR algorithm in window-limited online form were applied. Both methods succeeded in detecting a change in the vibration signals towards the end of the test. A companion paper submitted by the University of New South Wales outlines the corresponding diagnosis and prognosis algorithms applied to the vibration data.

Key Words: Generalized Likelihood Ratio, GLR, ARMA, Fault Detection, Bearing Fault, Epicyclic Gear Train, Vibration Analysis

1 INTRODUCTION

The vibration characteristics of epicyclic gear trains are more complex than fixed axis gears because the transmission path varies with the rotation of the planet carrier. Since epicyclic gear trains are common in helicopter transmission systems, it is important that suitable data sets be available for testing vibration based fault detection and diagnosis algorithms for helicopter transmission systems. The Helicopter Transmission Test Facility (HTTF) located at the Defence Science and Technology Organisation of Australia is a state of the art facility capable of testing a number of different helicopter transmissions systems at high loads under controlled conditions.

The HTTF was used to create vibration test data corresponding to the generation and propagation of a bearing fault in a Bell 206B-1 helicopter main rotor gearbox. In the first part of this paper, Section 2, the test procedure and outcome of the test is described in detail.

The remainder of the paper is devoted to the application of Generalised Likelihood Ratio (GLR) based fault detection algorithms to the vibration data obtained from the test. Before proceeding it is worthwhile introducing the various notations used throughout the paper. As sequences occur often, a compact subscript notation is used. The notation $y_{(j:k)}$ is used to represent the sequence $y_j, y_{j+1}, \ldots, y_k$ when $k>j$, and $y_j, y_{j-1}, \ldots, y_k$ when $k<j$. The convention of using uppercase letters for random variables and lowercase letters for the values of random variables is also adopted here. The symbol $\mathbb{R}$ is used to denote the set of Real numbers.

The general model used in statistical change detection is a (chronological) sequence of random variables or vectors, $\{Y_k\}$, $k=0,1,2,\ldots$, with known parametrised conditional density $p_\theta(y_k|y_{(k-1:1)})$. The parameter, θ, is constant (or nearly so) and equal to θ_0 before some unknown change time t_0 (i.e. for all $k<t_0$). At the change time, θ changes abruptly to a new value θ_1 where it remains for all time $k>t_0$. A typical change detection algorithm comprises a suitable test statistic, S_k, a decision function, g_k, which usually converts the test statistic into a form in which a suitable alarm threshold can be established, and an alarm time, t_a, to indicate when the decision function exceeds the chosen threshold.

As the problem formulation in statistical change detection is very general, the scope for applications is wide. Example applications include gain updating in adaptive tracking algorithms, signal segmentation for tasks such as speech recognition, and quality control (which includes fault detection). A comprehensive introduction to statistical change detection theory and applications can be found in the book [1].

The GLR is a powerful test statistic for use in change detection algorithms, however a well known problem with the online application of the GLR algorithm is that the number of operations to perform at each time step, k, grows with k to infinity. This difficulty comes about because the Likelihood ratio cannot be computed recursively. By restricting the algorithm to a data window of fixed size, the computational burden can be reduced to a manageable level, but the resulting *window-limited* algorithm becomes sub-optimal. This idea was first investigated in [2], in the context of detecting additive changes in linear state space systems.

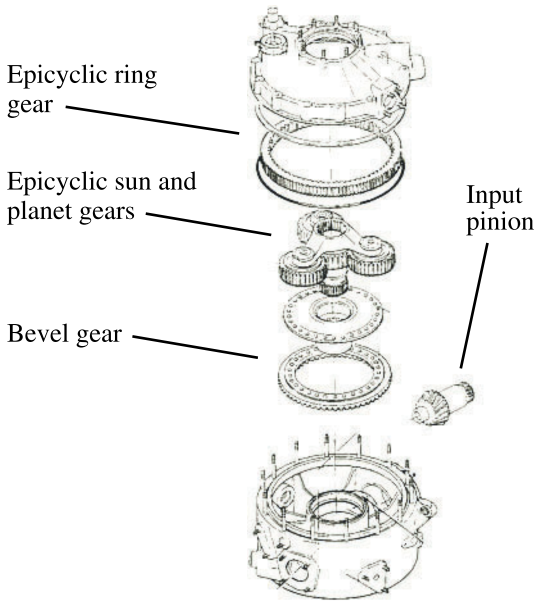


Figure 1: Major components of the Bell 206B main rotor gearbox.

It was shown in [3] that the performance of the window-limited GLR algorithm depends critically on the choice of window parameters, even in the simple case of detecting additive changes in white noise. The choice of window parameters also affects the choice of decision threshold which further complicates the issue. A number of authors [4, 5, 6] have addressed the issue of finding optimal window parameters for the case of additive changes in linear systems, however we are concerned here with the more difficult problem of detecting spectral (non-additive) changes in a signal, and in particular to spectral changes in an ARMA model.

In [3], a novel decision function for the window-limited GLR algorithm is introduced which is shown through simulations to be more robust against sub-optimal choices of window parameters. In the second part of this paper we compare the performance of the window-limited GLR algorithm with the new decision function to that of the window-limited GLR algorithm using the conventional decision function. In Section 3 the change detection problem is formulated in terms of the GLR algorithm, and in Section 4 the results are presented and discussed in detail.

2 TEST DESCRIPTION AND OUTCOME

The Bell 206B-1 main rotor gearbox, shown in Figure 1, comprises an input bevel and pinion, driven by the engine, and a single stage epicyclic reduction unit. The main rotor is driven through the planet carrier. There exist both three and four planet variants for the Bell 206 gearbox. The three planet variant was used for this test. Accelerometers were mounted on the gearbox casing in the vicinity of the pinion and bevel gear mesh and in the vicinity of the epicyclic ring gear. The input shaft also contained a tachometer producing one pulse per revolution of the input shaft. The tachometer signal was used to convert the vibration signals from time domain to angular domain.

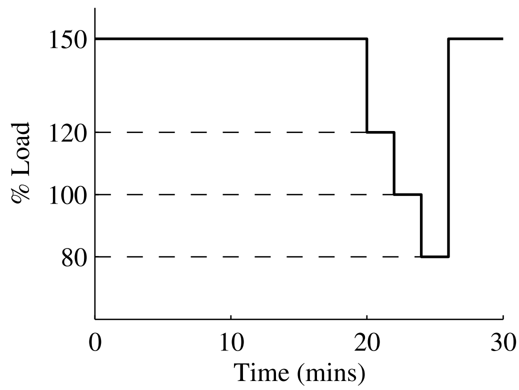


Figure 2: Gearbox load regime repeated throughout the test.

The aim of the test was to initiate and propagate a non-seeded fault in the gearbox. In order to achieve the aim in a reasonable amount of time, the gearbox was run at 150 percent of it's rated operating load, with periodic reductions in load to facilitate the recording of vibration signals. Figure 2 shows the load regime that was repeated every thirty minutes throughout the test.

The planet support bearing contains two rows of convex rollers and an outer race which is integral with the planet gear. The bearing sustained fatigue spalling and subsequent pitting damage to the inner race, outer race, and rolling elements. The majority of damage occurred to the inner race. This can be explained in part by the fact that the inner race does not rotate as the carrier rotates, therefore the load is concentrated in the same part of the inner race rather than being distributed around the whole inner race over time. Figure 3 shows the inner race of the planet bearing in which the pitting damage can be seen clearly in both the upper and lower tracks, with the majority of damage to the lower track.

3 PROBLEM FORMULATION

The vibration data from the test are modelled using Gaussian ARMA models. For positive integers p and q, the output of a scalar ARMA(p, q) process satisfies:

$$y_k = \sum_{i=1}^{p} a_i y_{k-1} - \sum_{j=1}^{q} b_j e_{k-j} + e_k \quad (1)$$

where y_k, a_i, b_j are real numbers. The *innovations*, $\{e_k\}$, are generated by a sequence of independent identically distributed (i.i.d) random variables with known density function. An important variable related to the innovations is the residual. The *residual* sequence of the ARMA process is defined by:

Figure 3: Inner race of the planet gear support bearing.

$$\varepsilon_k = y_k - \sum_{i=1}^{p} a_i y_{k-i} - \sum_{j=1}^{q} b_j \varepsilon_{k-j} \quad .(2)$$

For zero mean Gaussian distributed innovations, the conditional density for the ARMA process (1) is known to be:

$$p_\theta(y_k \mid y_{(k-1:1)}) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\varepsilon_k^2 / 2\sigma^2), \qquad \theta = (a_1, \ldots, a_p, b_1, \ldots, b_q, \sigma)^T, \qquad (3)$$

where σ^2 is the variance of the innovations, and θ is the change detection parameter.

A number of standard tests can be formulated for the ARMA process depending on the nature of the change expected. A different change detection algorithm exists for each test. To determine the correct test for this problem, an exploration of the behaviour of the parameter before and after the test is required.

Firstly, the magnitudes and phases of the frequency components of the gearbox vibration will be affected by the input and output loads applied to the gearbox, and on the operating temperature of the gearbox. In particular, changes in the gearbox casing temperature can cause changes in the casing response to excitations. The effects of load variation are minimised during the test by running the gearbox under constant load, and the effects of temperature variations are minimised by using only the vibration data recorded with the gearbox operating temperature within a fairly narrow range. Although these effects are minimal, they are not eliminated completely, therefore the parameter θ will have some variation with respect to time, caused by these variations both before and after the change time.

Before the change time, i.e. before the bearing fault is initiated, θ will be expected to remain approximately constant. But after the change time, as the test progresses, the condition of the bearing will deteriorate, therefore it is expected that θ will not be constant after the change. Furthermore the majority of the bearing damage occurs on the inner race, which does not rotate with respect to the planet carrier, therefore this damaged section remains in the load zone continuously. Thus the following reasonable assumptions are made concerning the formulation of the problem.

A1 There exists a (change) time, t_0, and a disjoint partition, Θ_0 and Θ_1, of $\mathbb{R}^{p+q+1}$ such that:

1. $\theta \in \Theta_0$ for all time $k < t_0$.
2. $\theta \in \Theta_1$ for all time $k \geq t_0$.

A2 The region Θ_0 is bounded and not too large.

Assumption A2 captures the idea that θ will remain approximately constant before the change. The notion of "not too large" shall be quantified later on in the development. Note that after the change time, the behaviour of the parameter is expected to change with time as the bearing deteriorates. This is in contrast to the assumptions used in the theoretical development of change detection algorithms, i.e. that the parameter be constant (or nearly so) before and after the change time. However, given that the behaviour of the parameter before the change is expected to be approximately constant (Assumption A2), and provided that Assumption A1 holds also, then the performance of the change detection algorithm will remain nearly optimal.

At this stage, we can say a little more about the regions Θ_0 and Θ_1. Given the nature of the bearing dynamics, it is most likely that the bearing fault will result in spectral changes in the vibration signature. Thus the change detection problem will be formulated as detecting a change in the $\{a_i\}$ and $\{b_j\}$ components of the Gaussian ARMA parameter, θ, subject to Assumptions A1 and A2.

3.1 The GLR Algorithms

The first algorithm, Algorithm 1, is the conventional window-limited GLR algorithm. For the problem statement consistent with Assumptions A1 and A2, the algorithm is given by:

$$\begin{aligned} t_a &= \min\{k : g(k) > h\} \\ g(k) &= \max\{S_j^k : k - n_1 \leq j \leq k - n_2\} \\ S_j^k &= \ln \frac{\sup_{\theta \in \Theta_1} \prod_{i=j}^{k} p_\theta(y_i | y_{(i-1:1)})}{\sup_{\theta \in \Theta_0} \prod_{i=j}^{k} p_\theta(y_i | y_{(i-1:1)})} \end{aligned} \tag{4}$$

where n_1 and n_2 are the window parameters, t_a is the alarm time indicating when the decision function, $g(k)$, exceeds the threshold parameter h, and $p_\theta(y_1|y_0) = p_\theta(y_1)$.

The second algorithm, Algorithm 2, differs from the first only in the decision function [3, (6.17)]:

$$g(k) = \max\{(g(k-1) + \bar{s}_k), 0\}, \qquad \bar{s}_k = \max\left\{\tfrac{1}{k-j+1} S_j^k : k - n_2 \leq j \leq k - n_1\right\}. \tag{5}$$

The log-Likelihood, S_j^k, requires an optimisation to be performed for each computation. It was shown in [3] that the local approach outlined in [1] could be used to obtain a local approximation to S_j^k which does not require any optimisations to be performed. The key components of the development are reproduced here.

Let $\theta_0 \in \Theta_0$ be the central, or average, value of θ before the change time. Let $\mathbf{I}(\theta_0)$ be the Fisher Information matrix, which for a Gaussian ARMA process can be written in block matrix form

$$\mathbf{I}(\theta_0) = \begin{pmatrix} \frac{1}{\sigma^2}\check{\mathbf{I}}(\theta_0) & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}. \tag{6}$$

An algorithm for computing the exact Fisher Information matrix for an ARMA process is given in [7], which can be used to compute the elements of $\check{\mathbf{I}}(\theta_0)$. Introduce the variables:

$$\alpha_{k-i} = -\frac{\partial \varepsilon_k}{\partial a_i}, \quad \beta_{k-i} = -\frac{\partial \varepsilon_k}{\partial b_i}, \quad \check{\mathbf{z}}_j^k(\theta_0) = \begin{pmatrix} \sum_{i=j}^{k} \frac{1}{\sigma^2} \check{\mathcal{A}}_{i-p}^{i-1} \varepsilon_i \\ \sum_{i=j}^{k} \frac{1}{\sigma^2} \check{\mathcal{B}}_{i-q}^{i-1} \varepsilon_i \end{pmatrix} \tag{7}$$

in which $\check{\mathcal{A}}_{i-p}^{i-1}$ and $\check{\mathcal{B}}_{i-q}^{i-1}$ are, respectively, vectors of α_k and β_k ordered backwards. Note that α_k and β_k are the outputs of AR models with the same AR coefficients:

$$\alpha_k = -\sum_{j=1}^{q} b_j \alpha_{k-j} + y_k, \qquad \beta_k = -\sum_{j=1}^{q} b_j \beta_{k-j} + \varepsilon_k. \tag{8}$$

The final variable, χ_j^k, comes about through a change of basis using the Kullback divergence as a metric. For a Gaussian ARMA process it is given by [1, §8.3.3]:

$$(\chi_j^k)^2 = \frac{1}{(k-j+1)^2}\left\{\sigma^2 \check{\mathbf{z}}_j^k(\theta_0)^T \check{\mathbf{I}}^{-1}(\theta_0) \check{\mathbf{z}}_j^k(\theta_0) + \frac{1}{2}\left[\sum_{i=j}^{k}\left(\frac{\varepsilon_i^2}{\sigma^2} - 1\right)\right]^2\right\}. \tag{9}$$

The equation for the log-Likelihood obtained using the local approach is [3, (6.47)]:

$$\frac{2}{k-j+1} S_j^k = \begin{cases} -(\chi_j^k - b)^2 & \text{for } \chi_j^k \leq a \\ -(\chi_j^k - b)^2 + (\chi_j^k - a) & \text{for } a < \chi_j^k < b \\ +(\chi_j^k - a)^2 & \text{for } \chi_j^k \geq b \end{cases} \tag{10}$$

where $a, b \in \mathbb{R}$ are the Kullback divergence parameters which quantify the regions Θ_1 and Θ_2 under the change of variable to χ_j^k. Note that a region of indecision is introduced, $a < \chi_j^k < b$, due to the local approximation.

4 RESULTS AND DISCUSSION

The data used in the application were taken from recordings with the gearbox operating at 100 percent of it's rated load. The first two data sets obtained from the test were used to fit ARMA models to the data. Using standard techniques of system identification [8], a Gaussian ARMA(13,1) model was found which approximated the pinion accelerometer data well. Similarly, a Gaussian ARMA(14,2) model was found for the ring gear accelerometer data.

The window parameters, n_1, n_2, were chosen based on the geometry of the gearbox. For simplicity n_1 and n_2 were set to the same value, n_1=n_2=n, and n was set to n=9000, which corresponds to approximately one rotation of the planet carrier. The Kullback divergence parameters, a, b, for the GLR algorithms were estimated using the first four data sets from the test. Specifically, χ_{k-n}^{k} was computed for 9000 points in each of the first four data sets, i.e. by sliding the data window across 9000 points in each data set. The parameter a was set to the maximum value of χ_{k-n}^{k} computed, and the parameter b was set to $a+0.25\sigma_{\chi}$, where σ_{χ}^{2} is the variance of χ_{k-n}^{k} computed on the first four data sets. The values of a and b obtained for the two accelerometers and two GLR algorithms are shown in Table 1.

Table 1.

Estimated Kullback parameters for the GLR algorithms.

	Pinion accelerometer		Ring gear accelerometer	
	Algorithm 1	Algorithm 2	Algorithm 1	Algorithm 2
a	1.4579	1.4961	1.2144	1.3187
b	1.4760	1.5151	1.2254	1.3379

The decision functions for Algorithm 1 are shown in Figure 4. The use of a finite window of past data means that the log-Likelihood, and hence the decision function for Algorithm 1, will have an negative expected value before the change and a positive expected value after the change. As can be seen in the plot of the decision function for the ring gear accelerometer, there is a change from negative values to positive average values occurring at approximately 145 test hours. This indicates that the algorithm is detecting the fault in the gearbox. There is no change to the negative values of the decision function for the pinion gear accelerometer. This indicates that the fault is not detected in this signal, which was to be expected since the pinion accelerometer is dominated by the pinion-bevel gear mesh which is unaffected by the condition of the planet gear bearing.

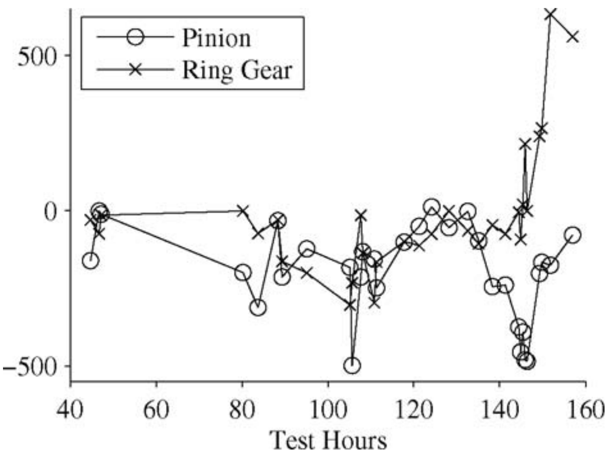


Figure 4: Algorithm 1: Window-limited GLR algorithm with conventional decision function.

The decision functions for Algorithm 2 are shown in Figure 5. The decision function in Algorithm 2 should be zero before a change and have a positive trend after a change. As can be seen in Figure 5, the decision function for the ring gear accelerometer detects a change at about 145 hours into the test, whereas no change is detected in the pinion accelerometer signal.

As mentioned in the introduction, the general formulation of the statistical change detection problem lends itself to a wide range of applications. Within a particular field of application there may be many different ways of implementing change detection algorithms. In the application given here for example, only minimal pre-processing of the raw data is done prior to implementing the change detection algorithm. The advantage of this approach is that the statistical change detection algorithm is being fully utilised to detect a wide range of changes in the signal, i.e. a wide range of faults. The disadvantage of the approach is the delay for detection. Both algorithms did not detect the change until the damage in the bearing had spread considerably. The danger here is in the possibility of a component failing before the algorithm responds.

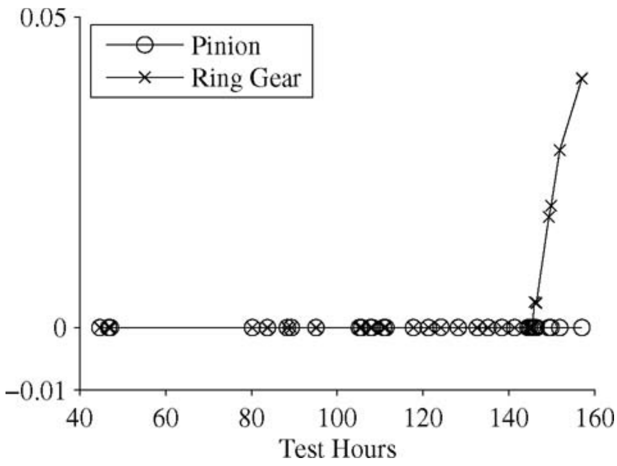


Figure 5: Algorithm 2: Window-limited GLR algorithm with modified decision function.

There are well known difficulties in detecting bearing faults. The vibration components from the (healthy) gears dominate the vibration signal masking the more subtle changes introduced by bearing faults. The level of difficulty increases further for planet gear bearings, where the motion of the bearing with respect to the transducer located on the casing introduces further modulation of the signal. Many of the current diagnostic algorithms for both gear and bearing faults rely on more complex pre-processing of the data to increase the signal to noise ratio for particular types of faults. For example, the dominant rotationally coherent gear meshing components can be filtered out using linear filters thereby increasing the signal to noise ratio of the bearing fault signature.

The general problem formulation in statistical change detection makes it easy to incorporate any amount of pre-processing prior to implementing the change detection algorithm, however there are disadvantages as well as advantages to be gained by using more complex pre-processing. For example, modelling specific fault types often leads to pre-processing algorithms which filter the data to enhance that part of the signal which the model predicts will be more sensitive to changes in the presence of that particular fault. This form of pre-processing can greatly restrict the range of detectable fault types.

5 REFERENCES

[1] Basseville M & Nikiforov IV (1993) *Detection of Abrupt Changes: Theory and Application.* Prentice Hall Information and System Sciences. Englewood Cliffs, New Jersey: PTR Prentice Hall.

[2] Willsky AS & Jones HL (1976) A generalized likelihood ratio approach to detection and estimation of jumps in linear systems. *IEEE Transactions on Automatic Control*, AC-21, 108–112.

[3] Galati FA (2005) *Statistical Signal Processing in Sensor Networks with Applications to Fault Detection in Helicopter Transmissions.* Ph.D. thesis, The University of Melbourne.

[4] Lai TL & Shan JZ (1999) Efficient recursive algorithms for detection of abrupt changes in signals and control systems. *IEEE Transactions on Automatic Control*, 44(5), 952–966.

[5] Nikiforov IV (1997) Two strategies in the problem of change detection and isolation. *IEEE Transactions on Information Theory*, 43(2), 770–776.

[6] Lai TL (2000) Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems. *IEEE Transactions on Information Theory*, 46(2), 595–608.

[7] Porat B & Friedlander B (1986) Computation of the exact information matrix of gaussian time series with stationary random components. *IEEE Transactions on Acoustic, Speech, and Signal Processing*, ASSP-34(1), 118–130.

[8] Ljung L (1999) *System Identification: Theory for the User.* Prentice Hall Information and System Sciences. Upper Saddle River, NJ: Prentice Hall PTR, 2nd edn.

IMPROVING THE INFORMATION GOVERNANCE OF PUBLIC UTILITIES THROUGH AN ORGANIZATIONAL KNOWLEDGE BASE

Stefano Gianella [a] and Willi Gujer [a]

[a] *Swiss Federal Institute for Environmental Science and Technology (EAWAG), and Swiss Federal Institute of Technology (ETH), Ueberlandstrasse 133, CH-8600 Dübendorf, Switzerland.*

Abstract: To develop and implement Information Governance (IG) in a public utility is a challenging task. While many studies focus on the "what" question of IG, less attention is paid to how IG may be realized within the utilities. However, well-organized IG is essential to guarantee effective management of information resources. We argue that IG is a complex and knowledge-intensive process. The ability to manage critical knowledge and to systematically support IG knowledge work are presented as key issues for public utilities. Most of their critical organizational knowledge is currently fragmented, not explicitly documented and difficult to share. We formulate requirements for designing IG for public utilities from a knowledge perspective. On the basis of three case studies of Swiss wastewater utilities, we present an approach for supporting IG via an Organizational Knowledge Base (OKB). We provide a framework for representing critical organizational knowledge for IG in an OKB, illustrate how an OKB supports knowledge work and explain why it guides public utilities towards more effective management of their information resources.

Key Words: Information Governance, Knowledge Management, Organizational Knowledge Base

1 INTRODUCTION

In today's increasingly competitive and cost-conscious operating environment, public utilities recognize the need to treat information as a valuable strategic asset. A large proportion of the administrative budget is typically spent on information-handling activities, e.g. for coordinating activities or for enabling data-intensive planning and design activities. Information is needed at all levels of a utility to support business functions and assist in attaining business aims and objectives. Making information available to the right worker at the right time and in the right place is vital for managing large and complex infrastructure systems.

In order to leverage the potential of information, public utilities have to configure their resources and capabilities to develop, exploit and maintain their information assets. This process is known as Information Governance (IG). IG is defined as a structure of relationships and activities designed to direct and control the information assets and the technology that supports them [1]. It establishes a framework of policies, procedures, standards and guidelines designed to ensure that the value of information assets is identified and exploited to the fullest extent.

The development and implementation of effective IG strategies and processes have been treated extensively in the literature over the past decade. Recent works have pointed out the importance not only of the "what" question of IG (e.g. improving links between information systems strategy and business strategy), but also of the "how" question, which includes the IG process and whether it yields effective outcomes [2]. Other researchers have pointed out that the process of formulating and implementing an IG strategy must be considered as a learning process for the organization that implements it [3].

This paper argues that IG is a knowledge-intensive and complex process for public utilities. The ability to manage critical knowledge for the IG process and to systematically support its knowledge work are presented as key issues for designing IG in public utilities. Although knowledge management has now gained wider acceptance, few utilities today are fully capable of developing and leveraging critical organizational knowledge to support their business functions. Many utilities are so complex that this knowledge is fragmented, difficult to locate and share, and therefore redundant, inconsistent, or not used at all. A dedicated method that explicitly considers knowledge management issues is proposed to support public utilities in designing their IG process. The paper presents design requirements for IG. It focuses on three case studies in public wastewater utilities. The findings from these studies are used to outline the proposed methodology.

2 A KNOWLEDGE PERSPECTIVE ON INFORMATION GOVERNANCE

2.1 IG as a complex and knowledge-intensive process

Complex and knowledge-intensive processes can be distinguished from other processes by a set of specific attributes and inherent problems. Several attributes for describing process complexity and knowledge intensity have been proposed by Eppler et al. [4]. The following inherent properties of the IG process can be highlighted on the basis of this proposal:

- Utilities often have no formal procedures designed to carry out the IG process, preferring to adopt ad-hoc approaches without clearly outlined activities. The outcome of IG then depends substantially on the people involved, and the internally available skills and capabilities of individuals as well as the knowledge and experience of external advisors play a major role. As a rule, it requires a lot of time to acquire such skills and capabilities. Because of today's fast-changing business environment and technological developments, knowledge soon becomes obsolete.
- IG is a process that consists of numerous activities involving many internal and external persons. Many interactions and communication schemes must be guaranteed in the context of problem-solving. Coordination mechanisms are generally complex and require time and effort. The required activities are highly interdependent and must consider the dynamic of environmental factors (such as policies for data standards).

Because of these particular characteristics, specific issues must be considered in the design of the IG process. These are intrinsically tied to the way critical knowledge for the IG process is gathered and assimilated both at individual level and at the collective level of a public utility. In the next section, we focus on two specific issues that we found to be decisive.

2.2 Key design issues for the IG process

As a complex and knowledge-intensive process, IG relies heavily on the experience and know-how of the people involved. Two critical issues thus have to be considered when designing the IG process within a public utility:

- *Which* knowledge is crucial for the success of IG and *how* can it be captured, managed and systematically exploited?
- *Which* knowledge activities are necessary for IG and *how* can they be supported?

2.2.1 Issue 1: Process knowledge of IG

Several models have been developed to describe the knowledge related to processes. Some examples can be found in Eppler et al. [4], Kwan et al. [5] and Stein et al. [6]. On the basis of these works, we define three types of knowledge that are appropriate for describing the critical knowledge relating to the IG process:

- **Knowledge about the IG process**: This first type of process knowledge encompasses all the skills necessary to design the IG process in a public utility. Specifically, this knowledge type answers the questions related to the process organization (responsibilities, stakeholder participation), the flow of activities, the necessary resources (time, expertise, money) and the process deliverables.
- **Knowledge generated within the IG process**: The second type of process knowledge includes the knowledge captured and generated in the course of the IG process. Two kinds of knowledge resources can be distinguished: knowledge that relates to the control of the process (e.g. goals and decisions) and knowledge that represents the outcomes of the process itself (e.g. business process analysis, information requirements, information inventory, information architecture).
- **Knowledge derived from the IG process**: This last type of knowledge consists of the insights gained from implementing IG activities. These include lessons learned, suggestions for improvement, experience with methods and tools and so on.

In order to capture, manage and exploit these different knowledge types, the problem of knowledge representation needs to be tackled. In particular, adequate ways of modeling the critical knowledge for the IG process as an object have to be provided. The main idea behind this "knowledge-as-object" approach is that it makes it possible to explicitly codify critical knowledge in the form of data sets [7]. Such object-like knowledge can be effectively captured, managed and exploited by means of information technologies in a similar way to data.

According to Zack [8], a meaningful representation of knowledge as an object must consider two basic aspects: *structure* and *content*. *Knowledge structures* provide the context for interpreting accumulated content. The basic structural element is the knowledge unit, a formally defined atomic packet of knowledge content that can be labeled, indexed, stored, retrieved and manipulated. Furthermore, a knowledge structure includes schemes for linking and cross-referencing knowledge units, e.g. for representing relationships. For its part, the *knowledge content* has to reflect the full range of critical knowledge that is needed. If we consider a specific process within an organization, a suitable knowledge representation should allow the embedded tacit and explicit knowledge of the process that generates and exploits it to be adequately and properly captured. In other words, the knowledge that is identified as critical for a process should be organized around the process itself [5]. In the search for an adequate representation, the fundamental question arises of determining which knowledge should be made explicit and which should be left tacit. Finally, it must also be stressed that not all the required knowledge can be explicitly codified in the form of data sets [6].

2.2.2 Issue 2: Knowledge work of IG

In order for IG to deal with its critical process knowledge, a set of embedded knowledge activities are required. These activities are intrinsically tied to each other and can be viewed as a support process within the IG process. The following activities are recognized as being suitable for describing the resulting knowledge work [6,9,10]:

- **Knowledge acquisition**: Knowledge that is not yet explicitly managed is identified and systematically captured. The source of this knowledge may be inside or outside the utility. It may already be documented or be present merely in people's heads. Knowledge that is captured has first to be verified and then integrated with existing knowledge.
- **Knowledge distribution**: Knowledge is made accessible for use. Ways of retrieving and accessing existing knowledge have to be provided to all knowledge workers inside and outside the utility.
- **Knowledge exploitation**: The exploitation of knowledge must be enabled for all knowledge workers within the context in which it is needed (e.g. within a business function). Knowledge content must be presented according to the requirements of the users.
- **Knowledge maintenance**: Knowledge has to be systematically updated, i.e. knowledge that is out of date has either to be removed or replaced.
- **Knowledge evaluation**: Knowledge has to be regularly evaluated and adapted to the changing requirements of its users.

Davenport et al. [11] demonstrated that improvement methods for knowledge work range from the classical top-down approach to a more "laissez-faire" philosophy that allows professionals to design and execute their own work. They found that an appropriate improvement method depends on several factors such as the organizational culture and the scope of the knowledge work. Lenk et al. [12] pointed out that an adequate approach to improving knowledge work in public organizations should focus on supporting knowledge workers. It has to emphasize the way people create, share and use knowledge resources and support these activities. Each knowledge worker should become aware of the knowledge work he/she carries out by doing his/her own job.

The role of information technology as an enabler for supporting and improving knowledge work has been broadly emphasized (e.g. Davenport et al. [11], Schwabe [10] and Zack [8]). Zack [8] suggests dividing knowledge management applications into two broad classes: *integrative* and *interactive* applications. Integrative applications, like IT-based knowledge bases, provide a means of integrating explicit knowledge in the organization. Interactive applications focus primarily on supporting interaction between people with tacit knowledge. Together, these approaches provide a broad set of knowledge-processing capabilities. However, a fundamental prerequisite for the effective use of information technologies to communicate and share knowledge in a utility is the existence of an interpretive context. This means that the structure and content of the knowledge have to be defined and communicated throughout the utility.

2.3 Current knowledge management practices of public utilities

Although knowledge management has gained wider acceptance both in the private and public sector, few utilities today are fully capable of developing and leveraging critical organizational knowledge to support their business functions. Nevertheless, developing effective ways of utilizing knowledge is recognized to be of increasing importance to enable utilities to offset competitive disadvantages and exploit latent corporate knowledge to its fullest potential (e.g. Blackwood et al. [13]).

Various IG projects of Swiss wastewater utilities provided us with insights into current knowledge management practices of public utilities:

- Knowledge is often fragmented, difficult to locate and share, and therefore redundant, inconsistent, or not used at all. Critical knowledge is usually unequally distributed among different knowledge carriers, and is mostly available only in tacit form.
- Utilities store knowledge in their procedures, norms, rules and forms. They accumulate such knowledge over time through learning from specific business issues. This knowledge is captured in routines and evolves in response to organizational experience. Knowledge therefore accumulates throughout the utility in the form of "knowledge silos". Individual knowledge is not systematically made accessible to others in the utility who need it. Different knowledge structures coexist in the utility and are not aligned with each other.

- Knowledge work mostly occurs implicitly at individual level. Well-defined knowledge activities, roles and responsibilities are not specified. The utilities are not fully aware of their knowledge resources.
- There is a lack of methods, techniques and tools for capturing knowledge from individuals throughout a utility. Instead, stand-alone applications are deployed independently in the utility. These applications (like GIS) can store a lot of the utility's critical knowledge. Valuable knowledge is therefore disseminated without any means of effectively integrating it.

These findings show us that current knowledge management practices in public utilities often fail to support basic knowledge processes such as capturing and securing existing knowledge, distributing knowledge and combining available knowledge. Good knowledge management involves the continuous streamlining of these knowledge processes to improve the learning capacity of the organization [14].

2.4 Design requirements for IG

By participating in IG projects of various Swiss wastewater utilities, the following requirements for an effective design of the IG process of public utilities have been identified:

- **Requirement 1: Manage critical process knowledge** – Knowledge about, within and from the IG process has to be systematically captured, exploited and updated. These knowledge resources have to be consistent, comprehensible, accessible and universally accepted in the utility.
- **Requirement 2: Make knowledge work visible** – The process of creating, distributing, exploiting, updating and evaluating knowledge about, within and from the IG process has to be clearly defined and embedded in the IG process.
- **Requirement 3: Use IT to leverage knowledge assets** – Knowledge about, within and from the IG process has to be formally structured and represented so that it can be stored in a repository. Activities such as storing, retrieving and manipulating knowledge objects must be supported by the use of IT.
- **Requirement 4: Define knowledge management roles** – Knowledge management roles and responsibilities within the IG process have to be explicitly assigned.
- **Requirement 5: Enhance knowledge awareness** – Individuals in a utility have to become fully aware of their knowledge resources and their importance for the whole utility.

In the next section, we present three case studies of Swiss wastewater utilities that show how these requirements can be implemented.

3 CASE STUDIES

3.1 Rationale

Between 2001 and 2004 we participated actively in three different IG projects of three public Swiss wastewater utilities (see Table 1). The rationale for conducting these projects was that all three utilities had recently made important investments in new information assets. The utilities became aware of the value of the collected information and recognized the importance of considering it as a key resource, together with people, finances and material assets. At the same time, the utilities realized that they faced a totally new challenge. In fact, an assessment of current IG practices revealed the following weaknesses:

- In some utilities (in particular some municipalities in case study 1) there is a complete lack of any recognizable IG process.
- Several utilities have still not considered important IG issues such as maintaining data quality, data integration, establishing data standards, metadata management and evolution of data models in response to changed requirements.
- In many utilities there is evidence that IG issues are recognized and need to be addressed. However, only ad-hoc approaches have so far been applied on a case-by-case basis.
- There is no global awareness in the utilities for the need to define appropriate standards and policies to govern the procedures and responsibilities for managing information. Responsibilities are left to individuals and are not clearly defined.
- In many utilities there is a lack of internal skills, capabilities and the experience necessary to manage information resources effectively.

Table 1

The three case studies presented

	1	2	3
Type of assets	WWTP and sewer network	Sewer network	Sewer network
Type of utility	Public joint association of 13 municipalities	Municipality	Municipality
Catchment area	76 km^2	51 km^2	48 km^2
Customers	53,000 inhabitants	38,000 inhabitants	52,000 inhabitants

WWTP: Wastewater Treatment Plant

3.2 Scope

The utilities considered here have recognized the need to improve their current IG practices. Throughout the projects, we concentrated our improvement efforts on three areas included within the scope of IG:

- **Management of information resources**. All the information resources of a utility need to be managed. IG must ensure that all these resources are known and that responsibilities for their management have been assigned.
- **Management of information processes**. All business processes give rise to operations involving one or other of the information resources of the utility. The processes of collecting, storing, accessing, modifying, deleting and archiving information must be properly controlled if the utility is to exercise satisfactory governance of its information resources.
- **Management of information standards and policies**. The utility will need to define standards and policies for its IG. Management policies will govern the procedures and responsibilities for managing information in the utility; technical policies and standards will apply to the IT infrastructure that supports the utility's information systems.

3.3 Methodological approach

Although agreement was reached on the need to improve current IG practices, an important issue remained open, namely: *how* do we proceed to bring about such improvement? For the three projects, we decided to adopt an approach based on seven steps (see Table 2).

Table 2

The seven steps of the approach used in the case studies

Step	Task	Goal
1.	Establish a participatory structure	Decision makers and key staff members participate actively in the project
2.	Achieve awareness for the *as-is* state	Understand how the utility currently manages its information assets
3.	Find agreement on what has to be improved	The major problems that need a solution are identified
4.	Define the *to-be* state	Define how the utility should manage its information assets
5.	Deduce and agree upon actions for improvement	An action plan is defined
6.	Implement the agreed actions	The agreed actions are implemented
7.	Evaluate the achieved results	Assessment of the impact of the implemented actions

In order to support each task, we developed a methodology that explicitly considers the design requirements of *Section 2.4*. The proposed methodology consists of three components that are explained below. In *section 3.4* we then show how the methodology was applied in the case studies and how the design requirements for IG were implemented.

3.3.1 Knowledge framework for IG

The first component of our methodology is a knowledge framework that we developed together with the public wastewater utilities considered here. This framework is used to properly structure and represent the embedded tacit and explicit knowledge of the IG process. Its structural design is based on six different perspectives (see Table 3). Each perspective emphasizes specific issues that have to be addressed within the IG process. Specific knowledge concepts are provided to represent the critical knowledge for each perspective.

3.3.2 Organizational knowledge base

We proceeded to construct a model containing these six perspectives on the basis of the identified knowledge concepts. In a first step, we used elements of existing enterprise models (e.g. Scheer [15]) to model each perspective. In a second step, we identified the relationships between the knowledge concepts used and integrated the models. We then refined the resulting integrated model by applying it to real-life utilities. The model was developed by using an object-oriented approach. Finally, we implemented the model in a relational database management system. We call the result of this implementation an *Organizational Knowledge Base* (OKB). It is the second component of the proposed methodology. This OKB can be used to carefully codify the critical knowledge for the IG process in the form of data sets and store it electronically in a database.

Table 3

The knowledge framework developed for this application

Perspective	Description	Related key issues	Knowledge concepts
Business processes	This perspective includes a description of business processes, business goals and business performance indicators. It describes the logical and chronological sequence of business and information-handling activities within the business processes. It also depicts the information requirements of these processes and the information produced or modified during the execution of business activities.	Which business processes within the utility have outcomes that are intrinsically dependent on information availability and quality? How will information be collected, stored, accessed, modified, deleted and archived within business processes?	• Goals • Performance indicators • Business processes • Business activities • Information handling activities • Information requirements
Information resources	This perspective includes a description of the necessary business data, their interrelationships, their ownership and the data models to use. It also describes the information objects that have to be provided for the business functions.	What information has to be developed, maintained and exploited to successfully achieve business objectives?	• Information objects • Business data • Data models • Data ownership
Organiza-tional structure	This perspective describes internal business units, external service providers and organizational structures. It specifies business roles and responsibilities of the business units. It also describes the operations of each business unit on business data within the execution of business processes.	Who in the utility will be responsible for collecting, storing, accessing, modifying, deleting and archiving information?	• Business units • Organizational structure • Business roles and responsibility • Data operations

Communi-cation	This perspective describes the information flows that take place with the execution of business processes, the business units and external service providers involved and the business data that is exchanged.	How will information be shared and distributed inside and outside the utility?	• Information flows
Information quality	This perspective describes the business rules, data quality rules and quality properties (such as accuracy, consistency, timeliness, integrity) required for the business data.	Which information quality level is necessary?	• Business rules • Data quality rules • Quality properties
Business events and triggered actions	This perspective describes the actions that are necessary to direct and control the information assets and specify the business events that will trigger them. Three types of actions are distinguished: data maintenance, data quality controls and data synchronization with external stakeholders.	How will the necessary information quality be maintained over time?	• Data maintenance • Data quality control • Data synchronization • Business events

3.3.3 Organizing and supporting the knowledge work of IG

We consider the OKB developed here as an enabling tool for organizing and supporting the knowledge work of IG. We have developed an architectural framework on its basis for designing the knowledge work of IG. This framework is illustrated in Figure 1. It is the third and last component of our methodological approach.

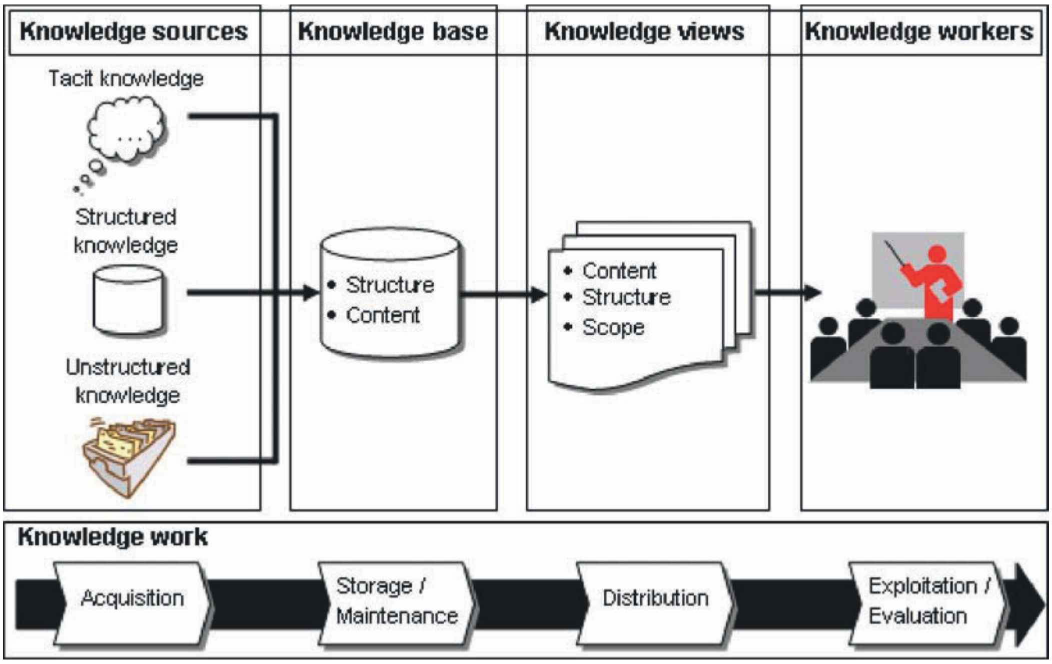


Figure 1. Architectural framework for supporting knowledge work.

The purpose of this architectural framework is to provide guidance for the utilities deploying the OKB. The knowledge base is the central component of the framework. Knowledge that is crucial for the IG process is captured from different knowledge sources and stored within the OKB. The OKB supports the maintenance of this knowledge. Retrieval of and access to existing knowledge in the OKB is provided to the participants of the IG process by means of knowledge views. Knowledge views can be seen as structured packets of knowledge derived from the OKB and supplied to address specific concerns within the IG process. The participants of the IG process are considered as knowledge workers who have specific knowledge requirements according to their roles and responsibilities within the process. They thus determine which knowledge views have to be configured for supporting their tasks within the process.

By implementing this framework in the utilities considered, we proceeded to develop a meaningful graphical user interface for accessing the OKB. We investigated the main concerns of the information managers and specified the key questions that had to be addressed. Basically, we found that the required knowledge views can be organized according to the different perspectives defined by the knowledge framework developed here (e.g. the business process perspective). A knowledge view can focus on one or more knowledge concepts (e.g. business data, business activities or information flows) and provide more or less details according to the descriptive attributes of the model. Several views can be obtained by combining different knowledge concepts using the specified relationships between them (e.g. business rules - business data). As a result of our implementation, we developed a set of functionalities to support the basic knowledge activities of the IG process such as storing, editing, deleting, searching and rendering knowledge. These functionalities are necessary to support the interaction of knowledge workers with the OKB.

In the next section, we illustrate how we applied our methodological approach in the three case studies.

3.4 Results

3.4.1 Step 1: Establish a participatory structure

A project team was set up in each case study. The resulting teams were composed of the information managers as well as the information administrators of the utilities. We insisted on including not only decision-makers or anyone else directly involved in the IG process, but also the key knowledge carriers of the utilities. In each case, we intervened directly in a coaching capacity, providing the team members with methodological know-how and data from other utilities.

The activities of the teams were organized as a series of workshops. We dedicated the first workshop exclusively to explaining the methodology we intended to adopt, to illustrate the activities we considered necessary and to explain the results we wanted to achieve. The first workshop aimed to define an agreement among the participants on the purpose of the project, project deliverables and the time frame for conducting the project. As a means of communication, we used the OKB filled with data from other utilities. By using specific knowledge views, we were able to provide the IG participants with real-life examples of the concerns we had to address. This allowed us to provide the participants with knowledge about the IG process very quickly. We observed that people initially tended to feel diffident and had different ideas on what had to be done. The discussions we had during the first workshop were mostly unstructured. We found that different participants normally considered and described the same problems, but from totally different perspectives (e.g. from a business perspective or from an information perspective). The knowledge framework we developed proved to be a suitable tool for disseminating a consistent description of important issues among all participants.

3.4.2 Step 2: Achieve awareness for the *as-is* state

After the first workshop, we proceeded to investigate the current information management practices of the utilities. This task normally required three to five workshops. These allowed us to collect the necessary knowledge from the available documentation and from the team members. Other staff members in the utility were involved to provide us with the missing knowledge as required. In these workshops, we worked interactively within the teams with the support of the OKB. We addressed the different perspectives of the knowledge framework in sequence and investigated each knowledge concept. We stored the captured knowledge directly in the OKB using the specifically designed graphical user interface.

The result of these workshops was a complete description of the *as-is* situation of each utility. Throughout the projects, the participants found it very helpful for developing a clear understanding of the way the utility operates. In fact, most participants only had access to that part of the knowledge that directly concerned their roles and responsibilities in the utility. None of them had direct access to the whole knowledge available. The workshops therefore represented a fruitful learning process for the participants and enabled considerable individual knowledge to be shared.

3.4.3 Step 3: Find agreement on what has to be improved

The documentation of the *as-is* situation of the utilities allowed us to identify several common areas where the current practices of managing information needed to be improved:

- Responsibilities are not defined clearly for all the information assets of the utilities
- Information-handling activities (e.g. collecting, storing, modifying, deleting or archiving data) are not explicitly defined within the business processes of the utilities
- There is no control of business events that produce or modify business data
- Data quality rules are not specified

- Information requirements are not clearly specified (often it is not clear if some business data still has to be managed)

While the utilities have adopted technical policies to develop information systems (e.g. for modeling data) or to exploit the IT infrastructure (e.g. IT security) in the past, management policies were still lacking. Therefore, in all three cases we agreed to develop suitable management policies for addressing the issues listed above.

3.4.4 Step 4: Define the *to-be* state

The first step towards defining management policies was to clearly specify a target state that had to be achieved. On the basis of a description of the *as-is* state we proceeded in sequence to define the *to-be* state for the utilities. Specifically, we organized two to three more workshops to interactively address the following issues:

- Specify information requirements and recognize the value of the available business data
- Specify the information objects and the business data that have to be managed
- Define responsibilities for the business data
- Explicitly define information-handling activities within the business processes
- Define roles and responsibilities of the organizational units involved in data operations within the business processes
- Define information flows
- Define a control mechanism for capturing business events that produce or modify business data and trigger the necessary actions
- Specify data quality rules

We used our OKB to support this step. For each issue, we first presented the *as-is* state by means of the knowledge views designed and then reviewed this briefly in the teams. This allowed us to discuss systematically how the *to-be* state had to look. The results of these discussions were directly recorded in the OKB. We found that the OKB was very useful in helping the project participants to specify the way things should have been done.

The knowledge framework proved to be an essential tool to design the *to-be* state consistently. It provided us with specific concepts and design rules to use for each issue that was addressed. This enabled us to effectively guide the project teams in the design process. Throughout the utilities, information managers found it to be an essential foundation for their management policies. In fact, it provided them with a straightforward way of dealing with a complex matter. Furthermore, the knowledge framework provided proved to be a suitable tool for communicating IG issues. They therefore decided to adopt this knowledge framework as a management policy for handling information.

3.4.5 Step 5: Deduce and agree upon actions for improvement

By comparing the *as-is* state with the *to-be* state of the utilities, we were able to deduce a set of actions necessary to realize the targeted state in each case. Examples of such actions are:

- Reengineering of business processes with a particular focus on information handling activities
- Making organizational units or external service providers accountable for the quality of the information produced
- Defining service-level agreements for information administrators (e.g. concerning data maintenance)

In each case we proposed an action plan and presented it to the project team. Emphasis was placed on the consequences for the utilities of implementing the planned actions. We also illustrated what costs could be expected and which benefits could be realized. The subsequent discussion with the project teams enabled us to improve the proposals. In a next step, priorities for each action could be defined. The project teams agreed upon the resulting actions plans.

3.4.6 Step 6: Implement the agreed actions

After having agreed upon the action plans, the utilities began to implement them. We observed that changing the way a utility operates and changing established structures is not a trivial process. We found that two critical success factors were especially important. Firstly, the commitment by the decision-makers who had to actively support the changes. Secondly, we found that communicating the changes throughout the utilities and making people aware of the need for change was essential. We observed that different levels of concern produced different rates of successful implementation. The system of participation that we built up during the workshops was very important to achieve the awareness and acceptance that current practices had to be improved. Making people at this early stage aware of change issues proved to be very important in making the changes succeed.

3.4.7 Step 7: Evaluate the achieved results

A successful change process must include steps for ongoing evaluation and reflection. We proposed to use the knowledge framework we designed as a checklist to investigate whether changes had really taken place and were of benefit to the utility. For example, in several utilities we observed a significant improvement in the data maintenance process. This can be explained as an effect of the better understanding of all the operations on business data that occur within a utility. In fact, by developing the knowledge framework, we precisely defined the activities in which business data was used, produced and modified. On this basis, we also defined the actions that had to be taken within the business processes in order to systematically update the data. As a consequence, it proved much easier to track data changes and to check whether they were effectively documented.

At the end of the projects, we conducted a review of the work done by the project teams in each utility. The project participants evaluated the way in which information management issues were structured as positive. The knowledge framework we developed was very helpful in focusing on specific issues and thus providing a clear understanding of the existing dependencies between the problems discussed. Our OKB and the knowledge views we provided were found to be suitable tools for communicating a complex matter. The workshops we conducted were also seen as a fruitful platform for sharing knowledge among different knowledge carriers in the utilities. The participants were able to learn from each other about the way things are done in the utility. This process was also important for becoming aware of the way things should be done. Finally, we investigated which problems resulted from using the knowledge framework for tackling IG issues. We found that the business process perspective presented the greatest difficulties to the project participants. This is because thinking in business process terms is still a challenge for many utilities.

4 DISCUSSION

On the basis of the findings from the three case studies, in this section we will review the design requirements for IG formulated in *Section 2.4*. We will describe how each requirement was implemented and discuss what effects we observed in the utilities considered.

Requirement 1: Manage critical process knowledge – We showed that knowledge about, within and from the IG process can be systematically managed through an OKB. We demonstrated that an OKB is a suitable tool for consistently integrating knowledge from different sources and making it accessible where it is needed throughout a utility. The OKB provides a utility with a consistent context for sharing and interpreting critical knowledge. We showed that the OKB is an appropriate tool for communicating issues of information management. It is very helpful to provide an IG team with knowledge about the IG process. We illustrated that the OKB supports different perspectives for addressing IG issues. This proved to be essential in view of the fact that the members of an IG team normally have different points of view and different concerns. We found that systematically capturing knowledge in the OKB helps an IG team to develop a clear understanding of the way the utility operates. This is extremely important for becoming aware of the concerns that IG has to address. We also demonstrated that the OKB can effectively guide an IG team through the process of designing the way things should be done. It therefore supports the specification of suitable management policies for the utilities. Finally, we emphasized that the OKB can also be used to capture insights from the IG process and can therefore stimulate the learning process of an IG team.

Requirement 2: Make knowledge work visible – We showed that an OKB enables knowledge activities within the IG process and make them explicit for the participants. The OKB is a suitable tool for building up a knowledge platform to support the acquisition, maintenance, distribution, exploitation and evaluation of critical knowledge. We organized these activities within the IG process in the form of workshops. This proved to be a suitable system for directly involving key knowledge carriers of a utility. We believe that making knowledge work visible enables the IG participants to reflect on the way the IG process is carried out and on the issues that have to be addressed. This shifts the attention of an IG team from the deliverables to the process that produces them. The process becomes more important than the results. We believe that this shift is essential for improving the effectiveness of the IG process over time. The IG process should not be regarded as static, but as a dynamic process that evolves over time as a learning system that continuously improves its effectiveness. We found that making knowledge work visible improves the learning of the participants and can catalyze the evolution of the IG process toward maturity.

Requirement 3: Use IT to leverage knowledge assets – We showed that knowledge about, within and from the IG process can be systematically represented by means of a knowledge framework. We illustrated how a knowledge framework can be used to implement an OKB, which can codify knowledge in the form of data sets. We showed that the basic knowledge activities can be supported by a set of IT functionalities. We found that IT can improve knowledge-processing capabilities for the IG process. This opens up new possibilities for analyzing complex issues of information management. The available knowledge can be fruitfully combined and provides new insights for solving complex issues. Codifying knowledge in a database opens up new ways of exploiting that knowledge and preserves it from losses. IT enables fragmented knowledge to be captured and integrated in a way that can provide information managers with the whole picture. This is necessary in order to achieve a clear understanding of critical issues and how they relate to each other. IT enables knowledge to be shared effectively in an IG team. It is essential to provide the decision-makers with the right knowledge.

Requirement 4: Define knowledge management roles – We showed that the OKB enables a new knowledge organization to be created for the IG process. Individuals are made responsible for documenting, sharing and maintaining critical knowledge for the whole utility. We found that this enabled a collective learning process within the IG process. IG participants could learn a lot from each other. We believe that this approach can improve the effectiveness of IG over time.

Requirement 5: Enhance knowledge awareness – We showed that an OKB improves the knowledge awareness of individuals in an IG team. Individuals became aware of their knowledge requirements and understood the importance of the knowledge they embodied for the whole utility. We believe that knowledge awareness is a fundamental condition for evolving the IG process over time.

5 CONCLUSIONS

On the basis of three case studies with Swiss wastewater utilities, we have proposed a methodological approach to supporting and improving current Information Governance (IG) practices through an Organizational Knowledge Base (OKB). We have developed a knowledge framework to systematically structure and represent the embedded tacit and explicit knowledge of the IG process. This framework was implemented in a relational database management system, thus providing us with an OKB for the IG process. On this basis we illustrated an architectural framework for supporting the knowledge work of IG. Finally, we outlined the benefits of the proposed methodological approach to the utilities considered.

In view of an increasingly competitive and cost-conscious operating environment, it is becoming more and more important to treat information as a strategic asset. Thus, public utilities should pay considerable attention to the design of their IG process and whether it yields effective outcomes. Because of the growing complexity of public utilities, managing information assets across business units and processes has become a challenging task. We argued that current knowledge-management practices fail to capture, manage and systematically exploit the critical organizational knowledge for an effective IG. We therefore emphasized that the IG process should not only be considered merely as a planning process for delivering an IS plan. Rather, IG is a learning process that requires continuous reflection on the way things are done. We believe that the approach presented here can stimulate the realization of IG as a learning process for the utility. It considers the participants of the IG process as knowledge workers and focuses primarily on the mechanisms allowing them to become fully aware of the utilities' needs.

We believe that the ability to manage critical knowledge and systematically support knowledge work is not only of crucial importance for IG, but also for other knowledge-intensive processes within a public utility (e.g. planning or auditing processes). We believe that our research may serve as an example for tackling this issue. The design requirements for IG presented here could be successfully applied to other processes. We have argued that knowledge about, within and from a process can be systematically represented in an OKB. In this regard, we illustrated how a comprehensible framework for representing knowledge as an object can be developed and implemented. Finally, we believe that the proposed architectural framework for supporting knowledge work could be fruitfully applied to other knowledge-intensive processes. A particular strength of this approach is that it focuses primarily on knowledge workers and their knowledge requirements for achieving their tasks. It emphasizes the way people create, share and use knowledge and improves the knowledge awareness of each individual.

6 REFERENCES

[1] ISACF. (2000) COBIT: Executive Summary. Rolling Meadows, IL: Information Systems Audit and Control Foundation.

[2] Grover V & Segars AH. (2005) An empirical evaluation of stages of strategic information systems planning: patterns of process design and effectiveness. Information & Management, 42 (5), 761-779.

[3] Huysman MH, Fischer SJ & Heng MSH. (1994) An organizational learning perspective on information systems planning. Journal of Strategic Information Systems, 3 (3), 165-177.

[4] Eppler MJ, Seifried PM & Röpnack A. (1999) Improving knowledge intensive processes through an enterprise knowledge medium. In J Prasad. (Ed) SIGCPR '99 Proceedings of the 1999 ACM SIGCPR Conference on Computer Personnel Research, New Orleans, pp. 222-230. New York, NY: ACM Press.

[5] Kwan MM & Balasubramanian P. (2003) KnowledgeScope: managing knowledge in context. Decision Support Systems, 35 (4), 467-486.

[6] Stein EW & Zwass V. (1995) Actualizing organizational memory with information systems. Information Systems Research, 6 (2), 85-117.

[7] Kakihara M & Sørensen C. (2002) Exploring knowledge emergence: From chaos to organizational knowledge. Journal for Global Information Technology Management, 5 (3), 48-66.

[8] Zack MH. (1999) Managing codified knowledge. Sloan Management Review, 40 (4), 45-58.

[9] Stein EW. (1995) Organizational memory: Review of concepts and recommendations for management. International Journal of Information Management, 15 (2), 17-32.

[10] Schwabe G. (1999) Understanding and Supporting Knowledge Management and Organizational Memory in a City Council. In HICSS '99 Proceedings of the Thirty-Second Annual Hawaii international Conference on System Sciences - Volume 1, p. 1075. Washington, DC: IEEE Computer Society.

[11] Davenport TH, Jarvenpaa SL & Beers MC. (1996) Improving knowledge work processes. Sloan management review, 37 (4), 53-65.

[12] Lenk K & Wengelowski P. (2002) Wissensmanagement für das Verwaltungshandeln. In T Edeling, W Jann & D Wagner. (Eds) Wissensmanagement in Politik und Verwaltung. Interdisziplinäre Organisations- und Verwaltungsforschung. Opladen: Leske+Budrich.

[13] Blackwood D, Gilmour D, Ball L, Dobson T & Taylor A. (2004) Knowledge retention and transfer in the operation of wastewater treatment plants. In NOVATECH 2004 Proceedings of Novatech 2004 Conference - Sustainable Techniques and Strategies in Urban Water Management - Volume 1, Lyon, pp. 473-480. Villeurbanne Cedex: GRAIE.

[14] Van Heijst G, van der Spek R & Kruizinga E. (1998) The lessons learned cycle. In UM Borghoff & R Pareschi. (Eds) Information Technology for knowledge management. Berlin et al. pp. 17-34: Springer.

[15] Scheer AW. (1998) ARIS - Modellierungsmethoden, Metamodelle, Anwendungen. Berlin et al.: Springer.

AN ASSET MANAGEMENT FRAMEWORK TO IMPROVE LONGER TERM RETURNS ON INVESTMENTS IN THE CAPITAL INTENSIVE INDUSTRIES

Kari Komonen, Helena Kortelainen and Minna Räikkonen

VTT (Technical Research Centre of Finland) P.O. Box 1000, FIN-02044 VTT, Finland.

Abstract: The general global problems in the capital intensive industry seem to be over-capacity and low returns on investment. The means to increase returns on investment are to decrease the operating costs or to increase the turnover of capital. From the physical assets' point of view, these requirements mean a need for dynamic and continual life cycle management, optimal capacity development, higher overall equipment effectiveness, higher reliability and flexibility of physical assets, and lower maintenance costs of production equipment.

In order to develop methods, such as dynamic life cycle management, optimal capacity development, OEE development etc., to manage the above mentioned problems a general asset management framework is needed. Theframework should be based on the business objectives of the firm and on analysing and modelling various businesses and business environments from the standpoint of the physical assets taking into account technological characteristics, economic structure and uncertainty in the industrial sector in question. In this paper, a methodology to build an asset management framework and a description how it results in asset strategy has been depicted and demonstrated.

Key Words: Strategic analyses, Asset management, Capacity management, Operation and maintenance

1 INTRODUCTION

The general global problems in the capital intensive industry seem to be over-capacity and low returns on investment (e.g. Siitonen 2003)[1]. The means to increase returns on investment are to decrease the operating costs or to increase the turnover of physical capital. From the physical assets point of view, these requirements mean a need for a dynamic and continual life cycle management, optimal capacity development, higher overall equipment effectiveness, higher reliability and flexibility of the physical assets, and lower maintenance costs of the production equipment. These trends have brought also the issue of availability warranties to commercial negotiations. Availability warranties are more demanding than e.g. traditional free-replacement warranties or replacements provided at a reduced cost to a buyer.

This development lays the cornerstones for 'asset management'. The expectations of various industries concerning 'asset management' vary: besides single techniques and viewpoints, comprehensive and integrated evaluation, development and optimisation techniques are also required. Some of the companies have seen the highest benefit from 'asset management' at a plant level, and some at a corporate level. The lack of methods and tools that combine technical assessment and economical considerations has also become evident. These facts require interdisciplinary know-how and competence (Komonen et al. 2005) [2].

In the capital-intensive industry, the operating time of the production equipment after the green-field investment is typically long. During the operating time numerous rebuilds, replacements and expansion investments take place. All of these decisions, together with the chosen maintenance strategies, affect the productivity of the physical capital. In order to meet the challenge of low returns on investment enterprises need to create an asset management strategy, which is based on constructed and systematic decision making at all the levels of the corporate hierarchy.

2 CONCEPT AND SCOPE OF ASSET MANAGEMENT

According to Mitchell (2002) [3] 'asset management' is ”a comprehensive, fully integrated, strategy, process, and culture directed at gaining greatest lifetime effectiveness, value, profitability, and return from production and manufacturing equipment assets”. In this approach, which was also adopted by VTT, two different aspects may be emphasized:

(1) Maintaining and improving the profit-making capability of production assets, and

(2) Maintaining and optimising the net asset value (physical assets)

'Asset management' could also include all of the decisions concerning physical assets: e.g. green-field investments, expansion and replacement investments, modernisations, improvements in Overall Equipment Efficiency (OEE), maintenance strategies and disinvestments. However, we made a decision not to include novel green-field investments e.g. for new products or for new markets. Today, when in many countries and companies the level of investments has decreased at the lower level than depreciations, the importance of 'asset management' of existing plants has become an issue in focus.

Greenfield investments have been based on assumptions concerning a longer term demand, competition, interest rate, technology development, cost level etc. Sometimes LCP-calculations have been exploited. A LCC/LCP analysis can be applied to the evaluation and optimization of the life cycle costs and profits of the investments taking into account specified performance, safety, reliability, maintainability and environmental requirements. Other reasons for performing an LCP-analysis can be, for example, to identify cost effective improvements and cost drivers (i.e. cost factors that have large-scale impacts on the investment profitability), or to choose between different suppliers and products. Various investment alternatives and their life cycle costs and profits can be evaluated and compared (Peltonen & al. 2002) [4]. However, as soon as the installation of the equipment has been carried out the business environment begins to change (Figure 1). Changes may occur in all the exogenous or internal factors which the investment calculations have been based on (Komonen & al 2005) [5]. The essential question present is how to sustain or improve the life cycle profits of the original investment. This is the core issue of 'asset management' of the existing physical assets. Some of the changes in the external and internal business environment call for operative and some of them strategic implications. Many strategic changes require a longer term perspective in decision making.

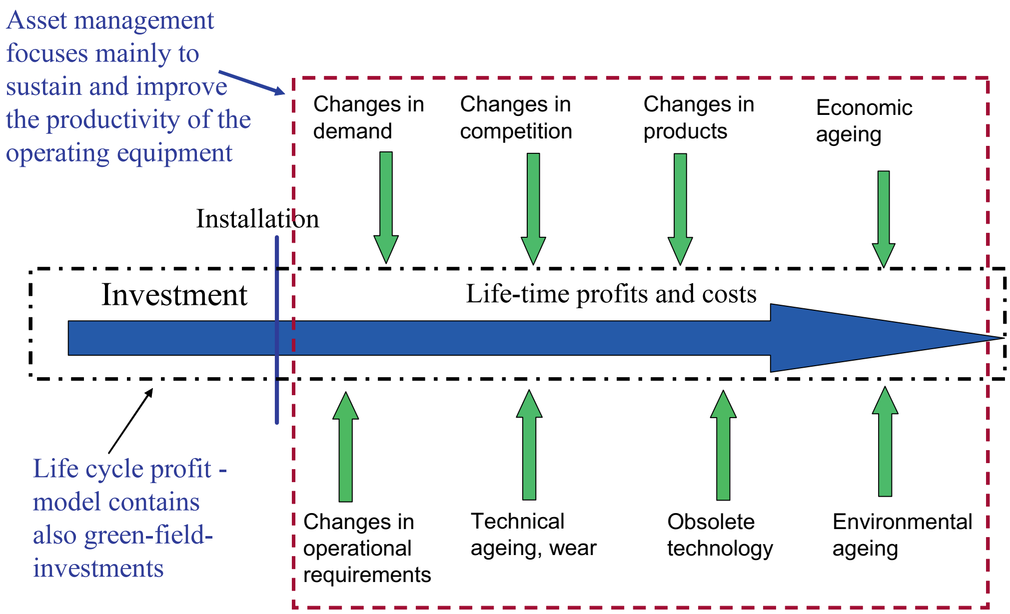


Figure 1. The focus of 'asset management

Discussions with the representatives of the capital intensive industry several issues concerning physical assets were brought on the stage. According to these discussions asset management may contain, ex ante, all the following issues:

- business requirements for physical assets and their use
- asset based capacity planning (analysis of real options)
- flexibility and agility management of production equipment

- dynamic continual life cycle optimisation (technical-economic view)
- maximisation of asset returns with respect to time horizon, risk carrying capacity and risk appetite
- physical asset centred risk management
- investment planning during the life time of the plant
- assessment of physical assets and of asset care systems
- determination of maintenance strategies
- capabilities, systems, plans and actions in order to sustain or develop the value, overall effectiveness and efficiency of physical assets and lower asset care costs.

However, although the above mentioned issues are obviously important part of 'asset management' as such they don't form an integrated and logical system to cover the whole field of asset management.

3 ASSET MAGAGEMENT FRAMEWORK

In order to develop methods to manage the above mentioned issues such as dynamic life cycle management, optimal capacity development, OEE development etc. the general asset management framework is needed. The framework is based on the business objectives of the firm and on analysing and modelling various businesses and business environments from the standpoint of the physical assets taking into account technological characteristics, economic structure and uncertainty in the industrial sector in question.

Corporate asset management requires systematic decision making at all the levels of the organisation. The proposed asset management decision making framework is schematically presented in Figure 2. The corporate asset strategy reflects corporate visions, values and mission, and the business objectives defined by the stakeholders and incorporates the information from the strategic analyses and scenarios. If the firm has several production facilities or plants the local asset related decision making needs to be in line with the corporate asset strategy and take into account the plant level objectives and constraints. At this stage, the technical and economic analyses and risk analyses create important input for the decision making. The plant or production asset strategy is allocated to the production systems and equipment and form a basis for the decision making at the shop-floor level.

The system and equipment level decision making is often regarded as 'asset management', but this standpoint is too limited to cover the identified needs of the industry. Using the definition introduced in Chapter 1 and the proposed asset management framework, the concept should cover the strategic and economic decision making and guide to the optimal use of the corporate assets to meet the business objectives.

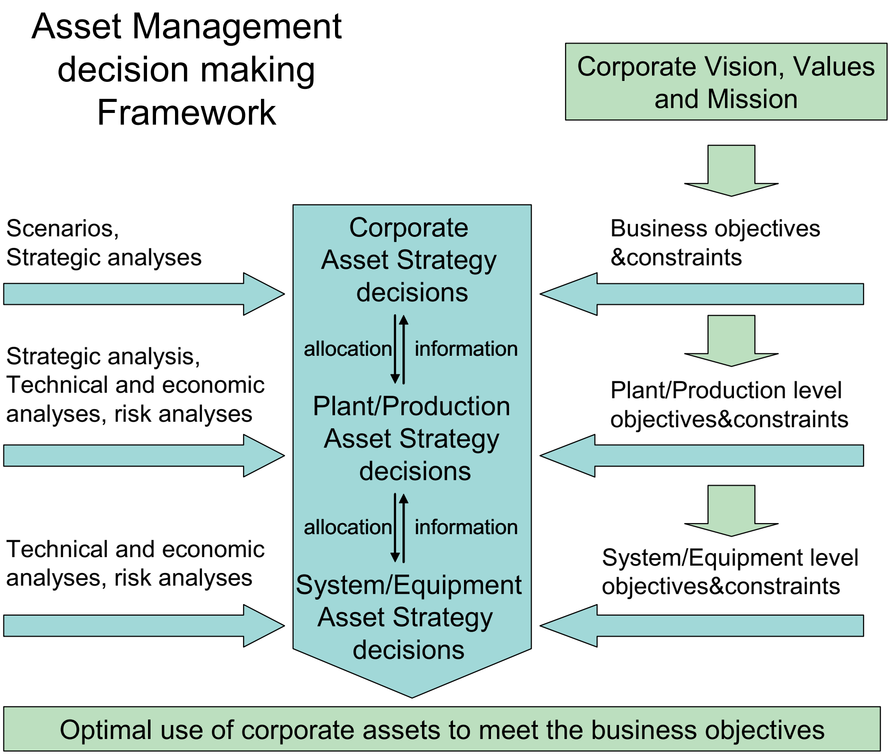


Figure 2. Schematic presentation of the Asset management decision making framework

4 DETERMINATION OF ASSET STRATEGY

Based on the above factors, analysing and decision making methods to determine the asset strategy have been developed for the various levels of the organisation as depicted in Figure 2. The scope and content of 'strategic asset management' instruments vary from level to level. The use various asset options, vertical integration, organisation of production system, mergers and choice of using asset services follow as a self-evident result of 'asset strategy'. The general development tendency of the business environment boosts strategically more flexible enterprises and production structures.

4.1 Corporate level

The corporate level is clearly strategic in nature and focus on the longer term perspective. The tools developed are close relatives to methods used in the area of 'strategic management'. The corporate asset strategy is based on the corporate business objectives and on the results of three main analyses: company, market and technology (Figure 3).

4.1.1 Markets

Market analyses determine the phase of the industrial development, asset related barriers to entries, uncertainty and volatility of industry structure and demand as well as sources of differentiation and regulatory acts by authorities. Market analyses comprise the basis for internal company analyses. The life cycle phase of the industry and the products in question have a remarkable impact on the flexibility requirements of the production system. The uncertainty and volatility of demand have the parallel effect on the flexibility requirements. An example of market analyses have been presented in Figure 4. The same type of classification, but concerning vertical integration, has been carried out by Harrigan (2003) [6].

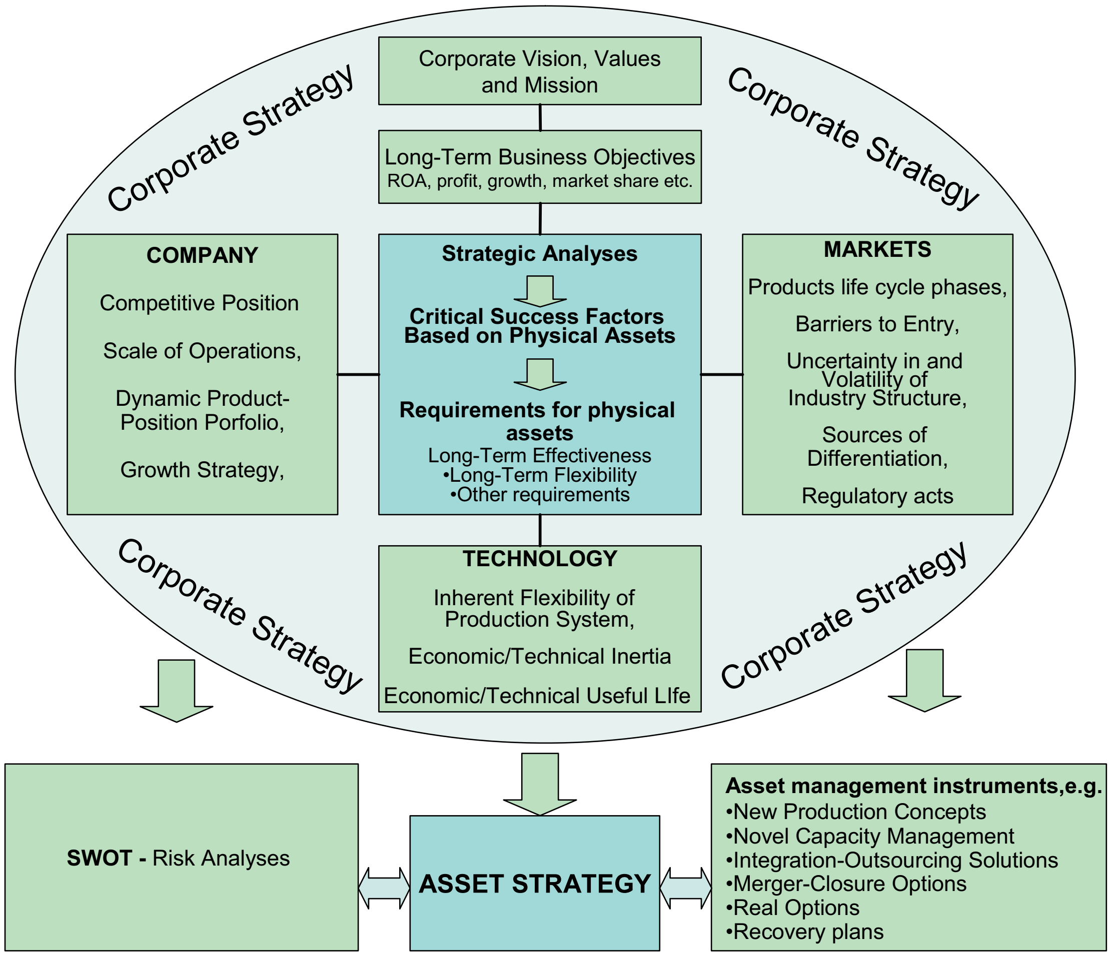


Figure 3. Schematic presentation of the Corporate Level Asset management decision making framework

4.1.2 Company

The analyses of the competitive position determine the portfolio of the competitive advantages in relation to physical assets. The scale of the operations influences possible longer run options to organise the production system in the optimal way in an effectiveness-flexibility co-ordinate. An example of the dynamic product-position portfolio is the map of company's product transitions in the life cycle- volatility co-ordinates. An example of the dynamic product-position portfolio is illustrated in Figure 4. The growth strategy describes company's actions to meet the growth objectives in the terms of the physical assets, outsourcing and mergers.

4.1.3 Technology

Technological factors have an essential impact on effective asset strategies. Specially, the flexibility of the production system (original investment), economic and technical inertia as well as the relative length of the economic and technical useful life of the equipment determine technological capabilities to meet requirements arising from the market-company interaction. The technological characteristics of the corporate production system determine significantly the possibilities of the company to meet flexibility or effectiveness requirements due to the market requirements.

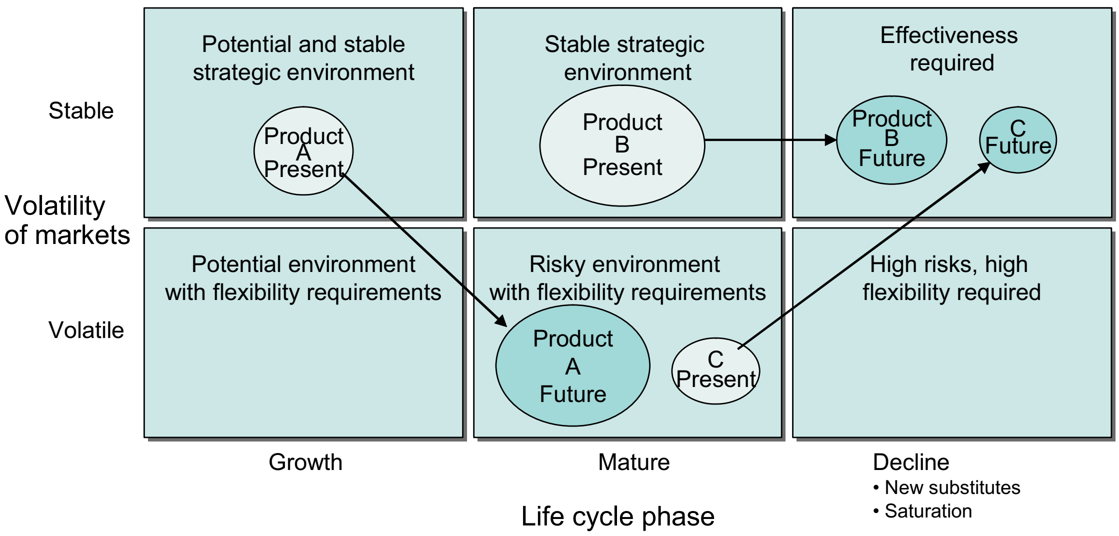


Figure 4. Example of product-position portfolio transition map in the selected time perspective

4.1.4 Strategic analyses

Company's internal strategic traits and development of the markets determine critical success factors from the physical assets point of view. The used technology gives limits and potential to the company's ability to meet the requirements of the markets and company traits. Based on the strategic analysis and critical success factors requirements for the physical assets can be identified. The asset strategy results from the requirements set by the markets and company's internal traits within the limits of technology in use and asset management options available.

4.1.5 Dynamic SWOT

In order to guarantee the long term quality of the asset strategy, risk analysis should be carried out. Dynamic SWOT analysis of the present and future circumstances could be used to improve the quality of the asset strategy and to decrease the uncertainty tied with the decisions made. The factors and analyses used in the determination of the asset strategy can also be used in dynamic SWOT-analysis.

4.1.6 Asset management instruments

Asset management instruments define a selection of strategic options in order to meet market requirements. At the same time they give the methods to find the optimal solutions to meet the corporate objectives. Examples of the asset management instruments are

- New Production Concepts
- Novel Capacity Management
- Integration-Outsourcing Solutions
- Merger-Closure Options
- Real Options
- Means to avoid risks and recovery plans

4.2 Plant level

The plant level is the mixture strategic and operative decision making in nature and focus on the middle and longer term perspective. The developed instruments are partly close relatives to methods used in the area of 'strategic management'. The plant level asset strategy is based on the plant business objectives and three main analytic dimensions: plant, plant business environment and technology of original investment (Figure 5). Strategic analyses based on the three above mentioned dimensions lead to the plant asset strategy and strategic choices. The strategic choices spring from the interaction of the possible options (asset management instruments) and asset strategy. The plant asset strategy and strategic choices give a framework to determine the optimal level of four performance factors: optimal capacity, operating rate, OEE (overall equipment effectiveness) and optimal maintenance costs.

Economic analyses are increasingly important in 'asset management'. The life cycle cost and profit objectives and life cycle cost structure have a significant influence on the plant asset strategy and strategic choices. LCP-analysis is a systematic approach aimed at assessing the total costs and profits of acquisitions, ownership, operating and maintenance, and disposal of an investment over its entire life-span. Within the asset management framework a challenge to meet is how to sustain or improve the life cycle profits of the original investment.

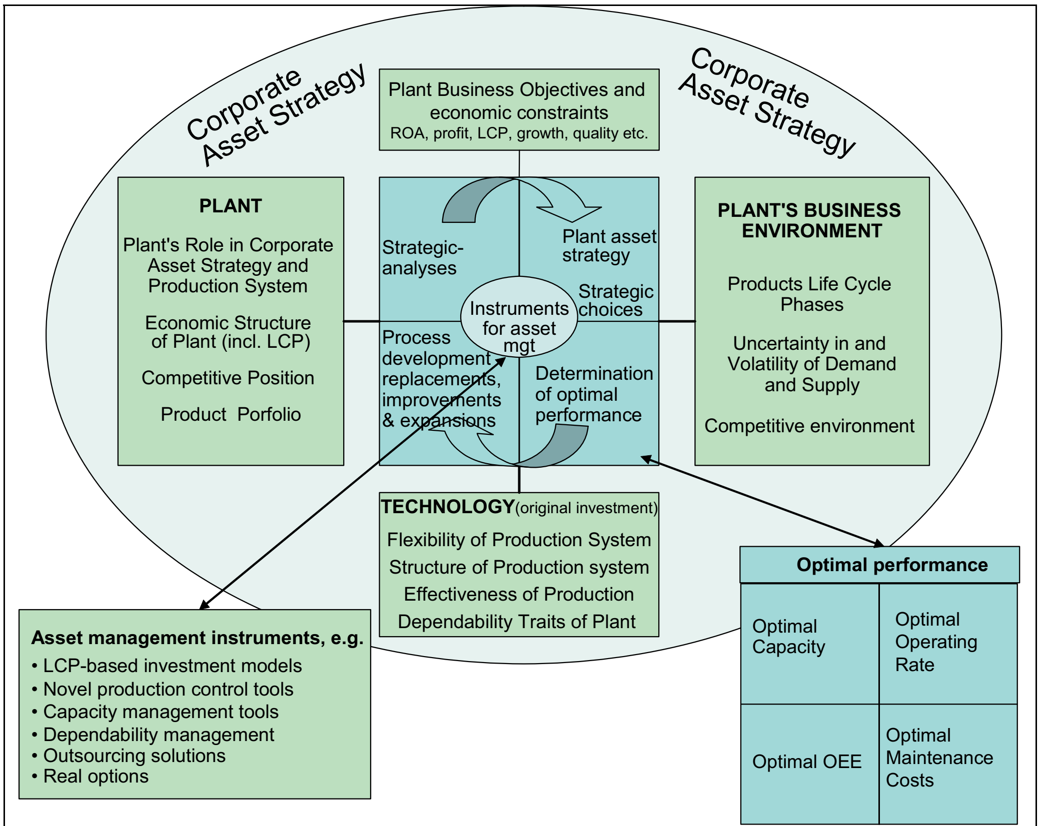


Figure 5. Schematic presentation of the plant level asset management decision making framework

4.2.1 Plant Business Environment

The analyses of the business environment reveal the phases of products life cycles, uncertainty and volatility of industry structure and nature of competition, for example, in terms of the sources of differentiation. The market analyses comprise the basis for the internal plant analyses. The life cycle phase of the industry and the products in question may have a significance impact on the flexibility requirements of the production system. The uncertainty and volatility of demand have a parallel effect on the flexibility requirements.

4.2.2 Plant

The corporate asset strategy determines the role of the plant in the corporate production system. The task of the plant could be the production of basic components, large scale production of one end product or small scale production of several products on the basis of customer order. The economic structure of the plant defines the significance of the physical assets in relation to the other variables. The life cycle cost structure determines profit potential related to various cost factors. The analyses of the competitive position determine the portfolio of competitive advantages in relation to physical assets. An example of the plant's dynamic product-position portfolio was shown earlier in Figure 4.

4.2.3 Technology of original investment

The flexibility of the production system affects capability to meet the volatility of the markets. The structure of the production system has a strong impact on the optimal activity level. The inherent effectiveness of the production system may give competitive advantage and with other traits influence the plant asset strategy. At the same time markets needs or plant's role in the overall production system determine the optimal position in the effectiveness-flexibility co-ordinate, and how well the used technologies match with requirements. The dependability characteristics of the plant differ from one technology to the other and have an impact on the optimal performance values. As depicted earlier, technological characteristics have an essential impact on the effective asset strategies. Specially, the inherent flexibility of production system, economic and technical inertia as well as the relative length of the economic and technical useful life of the equipment affects significantly the competitiveness of the plant.

4.2.4 Strategic analyses, strategic choices and performance

Plants' internal strategic traits and development in the markets determine critical success factors from the physical assets point of view. The used technology gives limits and potential to the plant's ability to meet the requirements of the markets. Based on the strategic analyses and critical success factors requirements for the physical assets can be identified. The determination of the asset strategy' and strategic choices result from the requirements set by the plant's business environment, company's internal traits, used technology and asset management options (instruments) available. The asset management options may include e.g. outsourcing, investments or better capacity management. As soon as the plant asset strategy and strategic choices have been decided, optimal performance objectives for capacity, operating rate, OEE and maintenance activities can be determined. The structure of the production system and other technological factors influence greatly the aspired performance level. Later in this paper some empirical evidence of this influence will be introduced.

The increase of capacity, operating rate or OEE requires e.g. investments, improvements, developing of maintenance processes and therefore costs money. The high values of performance figures are not necessarily optimal. The asset strategy, strategic choices and set optimal performance objectives give a basis for investment and improvement decisions.

4.2.5 Asset management instruments

Asset management instruments define the selection of various options in order to meet the requirements of the plant's business environment. At the same time they give methods to find the optimal solutions to meet the plant's objectives. Examples of the asset management instruments are following

- LCP-based investment models
- Novel production control tools
- Capacity management tools
- Dependability management and means to avoid risks
- Outsourcing solutions
- Real options

4.2.6 Empirical evidence

There exists empirical evidence of the impact of the business environment, plants internal traits and technological determinants e.g. on operating rate, OEE (overall equipment effectiveness) and maintenance costs (e.g. Komonen 2006) [7]. The structure of the production system has an impact on availability requirements as stated above. If a machine is the bottleneck of the production system, it is obvious that availability objectives are higher than in the case of a machine, which is not in the main production line and which has a low operating rate (Komonen 2002) [8]. "An indicator to model almost any production system is the degree of integration (plant replacement value divided by the number of front line production operatives in day shift), which is a multipurpose meter to measure mechanisation rate, continuity of processes and scale economies. Generally speaking, we can state that e.g. in the chemical industry and in the pulp and paper industry the integration level is high and in the manufacture of metal products it is lower. The impact of degree of integration on the above mentioned performance indicators has been demonstrated in Table 1: the higher the degree of integration of the production systems the higher availability performance they show. The same applies to the planning rate of the maintenance activities" (Komonen 2006) [7].

Table 1

The overall equipment efficiency (OEE), availability, proportion of planned and scheduled maintenance and proportion of preventive maintenance of production system as a function of integration level in Finland in 2000-2003

(N =261; N= 85-89 in each category)

Integration level	<0,65	0,65-4,5	<4,5
OEE %	69,5	72,7	80,3
Availability %	88,3	87,3	93,4
Proportion of planned and scheduled maintenance %	55,0	68,6	72,2
Proportion of preventive maintenance %	31,8	36,3	41,3

In order to make a more specific examination, multivariate regression analyses were carried out branch-wise using such exogenous independent variables as integration level, operating rate, plant size and technical capital intensity. Exogenous independent variables are factors which are often in the short run beyond the scope of production and maintenance managers to determine, but which have considerable impact on the objectives, e.g. availability or maintenance costs. The results of statistical investigations for three industrial sectors are presented below. The examples presented here are the chemical industry, processing of food and the combined data of processing of food, chemical industry and manufacture of metals. The number of plants in each sample varies, since some plants were not able to answer all the questions in the data collection form.

In the three above mentioned industrial sectors (Table 2), the degree of integration only had a statistically very significant influence on OEE. The impact of operating rate was nearly statistically significant. The third exogenous variable having some impact on OEE level was "Plant turnover / plant replacement value" (technical capital intensity). The determining power of the model was high (adjusted R^2=0,602. The statistical significance of the result was also very high ($p<0,0000$). In the chemical industry, the degree of integration only had a meaningful influence on OEE. However, its determining power was very high (adjusted R^2=0,774). The statistical significance of the results was also very high ($p<0,0001$).

Table 2

The variables having an impact on OEE in the processing of food, chemical industry and manufacture of metals

Dependent variable: OEE (Overall equipment effectiveness)

Adjusted R^2= ,602	F(3,29)=17,137	p<,0000
	t(29)	p-level
Intercept	6,566	0,000
Degree of integration	3,263	0,003
Operating rate	1,818	0,079
Plant turnover / plant replacement value	1,259	0,218

The relationship between OEE and integration level in the chemical industry is demonstrated in Scatter diagram 6 and the results for the combined data of three industrial sectors (processing of food, chemical industry and manufacture of metals) has been illustrated in Figure 7. In the scatter diagrams, predicted and observed values equal on the diagonal. The cases above the diagonal represent better performance than expected. In many other industries the degree of integration had also the most significant effect on OEE or availability of production equipment.

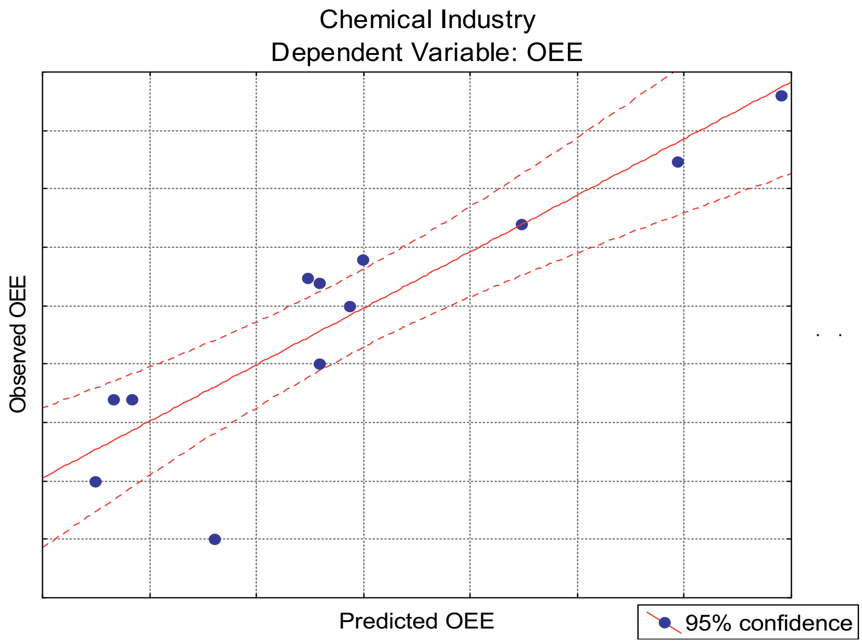


Figure 6. The predicting power of integration level on OEE in the chemical industry (Komonen 2006) [7]

As another example, it was interesting to notice that e.g. in the processing of food and liquid, the variables explaining maintenance costs were partly parallel with the variables explaining variation in OEE, although the direction of the influence was not always the same. However, in this case the replacement value was the most significant indicator. The degree of integration was the second statistically significant independent variable. The relationship between 'maintenance costs in relation to plant replacement value' and exogenous variables is presented in Table 3 and graphically demonstrated in Scatter diagram 8. In the scatter diagram, predicted and observed values equal on the diagonal (Komonen 2006) [7].

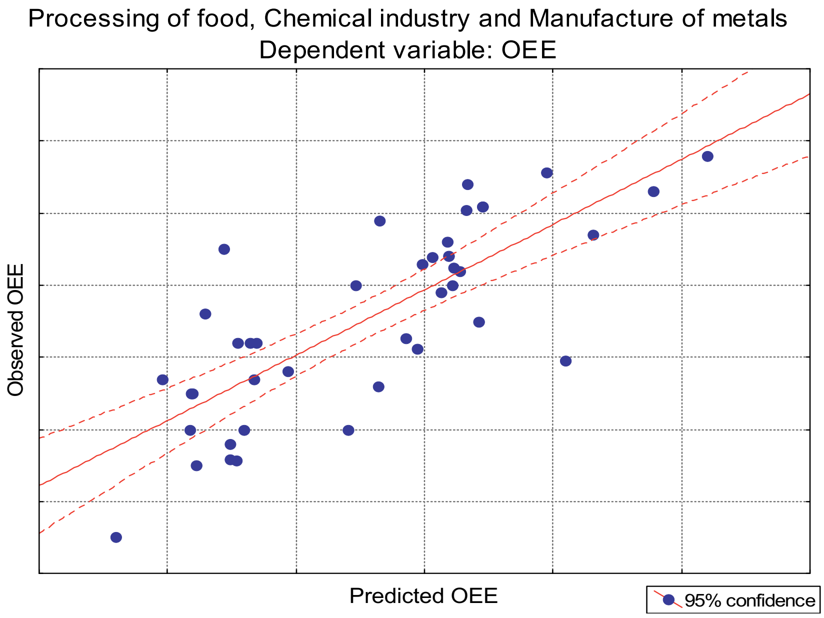


Figure 7. The predicting power of exogenous variables in the compound data of three industries

Table3

The variables having an impact on relative maintenance costs in the processing of food and liquid

Processing of food and liquid (N=34)

Dependent variable: Production equipment, Maintenance costs / Replacement value

Adjusted R^2= ,542	F(4,29)=10,773	p<,00002
	t(29)	p-level
Intercept	3,865	0,0006
Replacement value	-4,386	0,00014
Degree of integration	-2,275	0,030
Plant turnover / plant replacement value	1,859	0,073

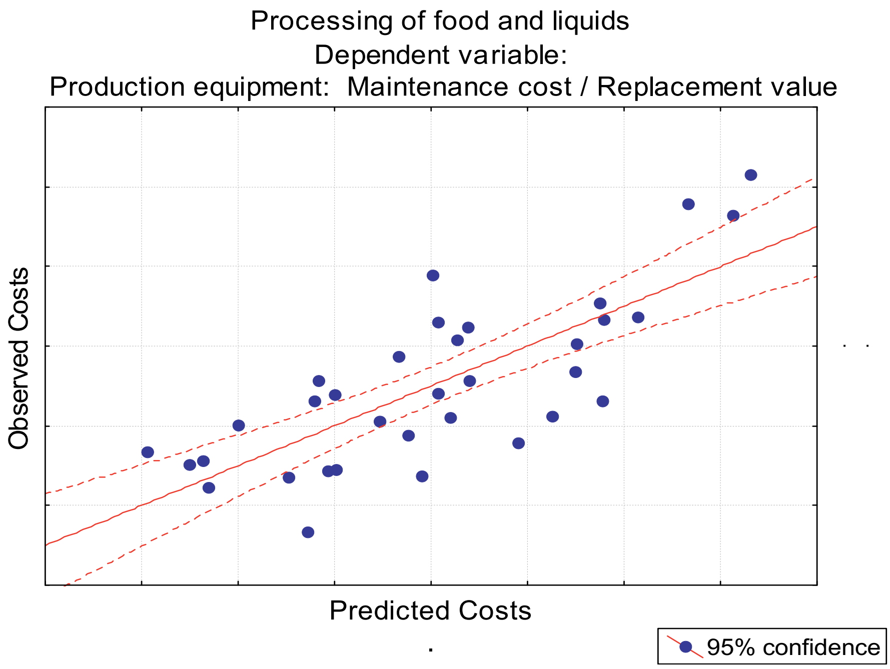


Figure 8. The predicting power of exogenous variables on relative maintenance costs in the processing of food and liquid (Komonen 2006) [7].

4.3 System, equipment and item level

The focus of the equipment level asset management is on the maintenance function. Surely, the effect of the operation of the production processes has a great role in achieving high performance level. However, in this presentation the latter perspective has a minor role. The life cycle management of machines and their components are, of course, important from the machine performance point of view. Still, the justification of the machine level investments springs up from the plant level requirements. Thus, as a summary, in this presentation the machine level maintenance is a synonym for 'machine level asset management'.

The basis for the machine level examination in presented in Figure 9. The description of the maintenance system starts with the business requirements of the plant in question. These requirements have been summarized in the list located in the upper left box of the diagram. The plant asset strategy, customer satisfaction and safety considerations guide also maintenance operations. Requirements for the maintenance function and maintenance objectives spring from the above mentioned business requirements.

The system of the maintenance function consists of four separate sub-processes

- maintenance planning
- resources management and development
- management of maintenance processes
- execution
- follow-up and continuous improvement

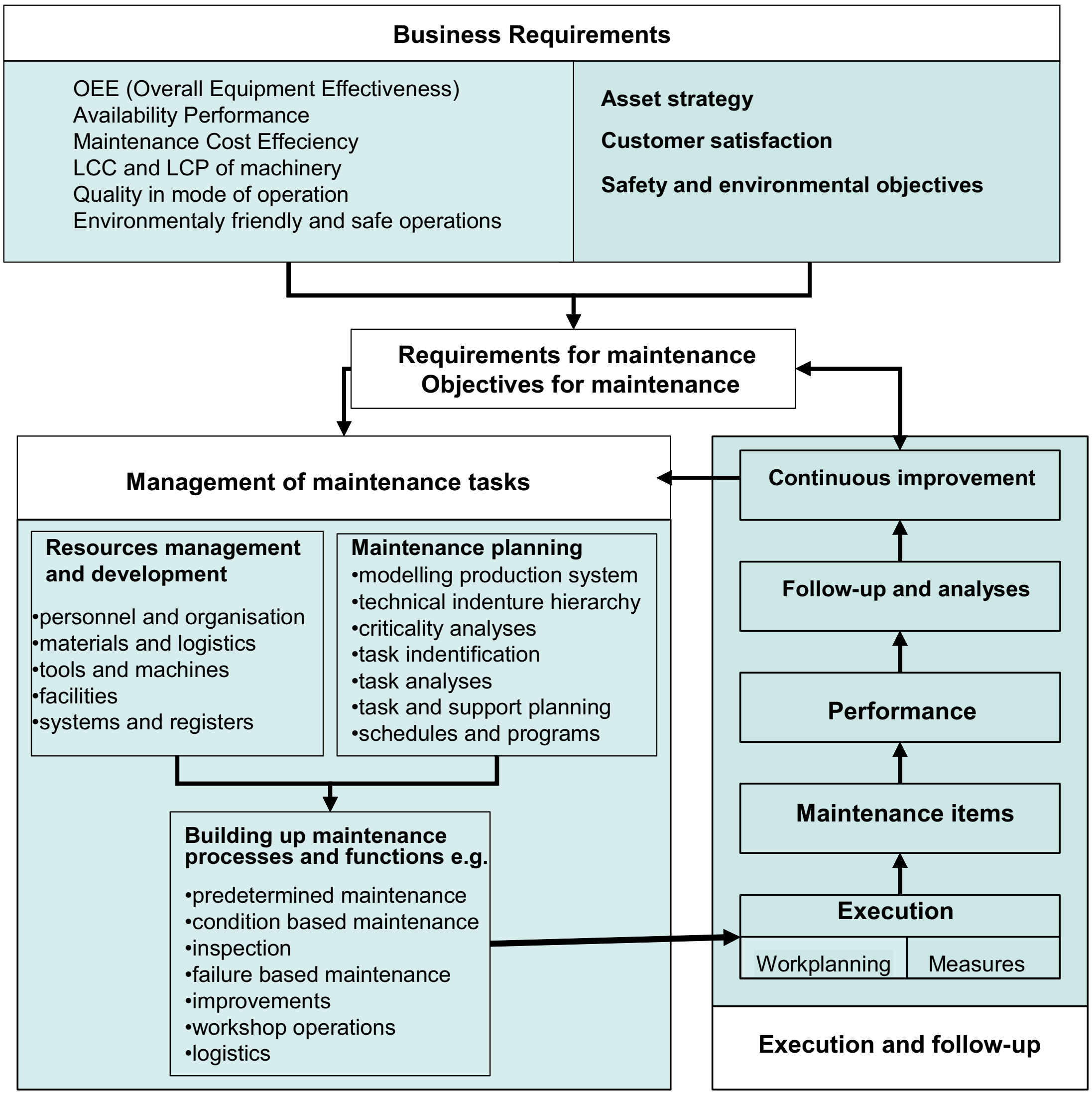


Figure 9. The system of the maintenance function

Maintenance planning includes e.g. modelling of production system in order to build up machine hierarchy for criticality analysis and maintenance policy. Maintenance planning continues with task identification, task analyses and task and support planning. The following step is to determine schedules e.g. for preventive maintenance and work lists for planned shutdowns. 'Resources management' is composed of planning, design, procurement, hiring, maintenance and development of maintenance resources. Organisation, IT-systems and condition monitoring equipment are essential parts of the maintenance systems. The third theme, maintenance processes depict the way, how the maintenance planning and execution are planned to be carried out and how the resources are planned to be used. The execution phase consists of work-planning and actions taken by the maintenance crew. The execution of maintenance actions leads to the performance of the machinery. The result of the maintenance system can be analysed and as a conclusion the improvements of the maintenance function can be carried out. The essential steering tool of the maintenance function is the system of performance indicators, which may cover business-oriented economic dimensions as well as organisational and technical factors.

There exist a lot of well known tools to carry out maintenance planning such as RCM, instead, the methods to develop maintenance processes are less frequent. Some writers, such as Kelly (1998) [9], are exceptions and have contributed also this research area.

There are various kinds of views, approaches and tools to develop and manage the maintenance function. In the field of the maintenance management and research eg. following categories can be identified:

1. overall approaches and models for maintenance management
2. asset centred planning tools and models, and statistical tools of reliability engineering
3. people and process centred methods, quality and organisational tools
4. process centred, work-planning approaches
5. mathematical economic and technical optimisation models
6. empirical research concerning causal relationships between various factors
7. continuous improvement

Some of these approaches have been discussed below.

(1) Overall maintenance management models include LCP-models, 'asset management', terotechnology and other holistic approaches on the maintenance function. LCP-models (life cycle profit) are quite a popular topic in research and consulting, but in reality they have been applied less than it could have been expected. Too many of the papers on LCP-models are listing the cost elements of LCC. However, the principle problems in LCC –calculations are the empirical responses of the changes in parameter values. 'Asset management' approaches extend the maintenance management view e.g. into choosing what equipment should be bought and when. Terotechnology emphasises the use of feedback from the field in the product development and design. Thus, there are connections between terotechnology and “life cycle profit” approaches. The basis for the good management is the system of performance indicators, which gives an instrument to identify the result , success and development needs of the maintenance function (Komonen 2004) [10].

(2) Asset centred planning tools include a wide range of methods and techniques, which are well known in the field of maintenance but too seldom used in the practical life. Examples of these methods are e.g. Quick Analyses of Production System (QAPS, developed in VTT), RCM, Experience Based RCM (EBRCM developed in VTT), RBI (Risk Based Inspection), FMECA, Fault Tree (FTA) and many other applications. RCM (Reliability Centred Maintenance) is a typical example of asset centred planning models. RCM is actually *a combination of good practices and methods*. Surely, many plants and maintenance managers have used some of the methods in RCM-toolkit. Mathematical models and statistical tools of reliability engineering can be used to support equipment and item level asset management.

(3) The third major approach contains people, organisation and process centred management tools. Perhaps, the best known concept in this field is 'total productive maintenance', TPM (Nakajima 1989) [11]. However, there are many other tools such as the use of maintenance process descriptions or process development implemented by group-work methods, simulation game based approaches or analyses of good working practices. 'Total productive maintenance emphasises the overall equipment effectiveness, operator centred maintenance, co-operation between various functions such as operators and maintenance technicians, adoption of proper working methods and attitudes, predictive and preventive maintenance methods. In this review, TPM has been interpreted to be a people, organisational and process centred management instrument, although it owes some features, which could be addressed partly to the asset centred category.

Maintenance processes to be determined, designed and developed may be grouped in several ways. Below, some examples of maintenance processes and functions have been presented:

- predetermined maintenance
- condition based maintenance
- inspections
- failure based maintenance
- improvements

- major shutdowns
- workshop operations
- materials function
- operator maintenance
- etc.

(4) Work-planning converts maintenance analyses, resources and designed maintenance processes into actions. CMMS tools and individual modes of operations guide and support the execution of "daily maintenance". E.g. Kelly (1998) [9] has approached maintenance planning from the standpoint of daily management, work planning and control.

(5) Economic and technical optimisation models have been of great interest to the practitioners of operational research. They treat such topics as the determination of optimal reliability investments (e.g. condition monitoring devices and programmes), optimal warranty policies, optimal maintenance policies, optimal timing of maintenance actions and optimal size of repair crew, optimal selection and number of spare parts. The search for optimal maintenance policies has generated e.g. a large literature of replacement models. These models could be easily used in practice. However, this is obviously not done to a great extent (Scarf 1997) [12]. In the field of warranty cost analysis W,R Blischke and D, N, P Murthy (1994, 1996) [13,14] have produced comprehensive and versatile handbooks, which present various aspects of warranties including optimisation models. Although the issues of a product warranty include several more important aspects than maintenance related issues, the maintenance function is an important tool to control the risks attached to product warranty and especially availability guarantees.

(6) An important neglected subject matter is the empirical research of the maintenance function and 'asset management'. It is very difficult to proceed in the theory formation and in the development of practical but scientifically grounded tools, if our knowledge on the “natural laws” of maintenance is not increasing. Benchmarking has been seen among maintenance managers and specialists as one of the major tools for development. However, benchmarking is not a possible option without comprehensive empirical research. There exists very little empirical work on the causalities in the maintenance field. Refreshing exception are the papers of Laura Swanson (1997 and 2001) [15, 16] and V.N. Bhat (2000) [17]. In Finland, Komonen has published papers and books concerning empirical relations between various performance indicators (e.g. 1998, 2002, 2006) [18, 8, 7].

(7) Continuous improvement is nowadays one of the critical success factors of any production plant. Several good principles of improvements could be listed. For example, active use of the versatile system of performance indicators, active and careful use of CMMS, participative leadership culture, resources for to carry out development activities, well-working cooperation between operations and maintenance function, optimal use condition based maintenance and inspections, viable quality system and systematic use of internal and external benchmarking systems.

4.3.1 Economic focus within various approaches

LCP-models, 'asset management' and market models are by their basic nature economic models. The same applies to mathematical optimisation models. RCM and TPM approaches include economic consideration, but they are not explicit quantitative models. For example, although TPM focuses on OEE, it does not employ profit or cost considerations. Some models and tools of reliability engineering employ qualitative economic considerations, usually in the form of risk assessment. Empirical studies focusing on economic causalities are surprisingly rare.

5 CONCLUSIONS

In the VTT's approach to 'asset management' two different aspects have been emphasized:

(1) Maintaining and improving the profit-making capability of production assets, and

(2) Maintaining and optimising the net asset value (physical assets)

This concept means that 'asset management' is a broader concept than just a business driven maintenance. 'Asset management' could also include all the decisions concerning physical assets: e.g. expansion and replacement investments, modernisations, improvements concerning overall equipment effectiveness (OEE), maintenance strategies and disinvestments. Also the options which offer other solutions than the exploitation of own equipment such as outsourcing, mergers, leasing belong to the field of 'asset management'. Today, when in many countries and companies the level of investments has decreased at the lower level than depreciations, the importance of 'asset management' of existing plants has become an issue in focus.

The asset management framework was divided into three levels: corporate, plant and equipment level. We have noticed that the equipment level has been quite well covered. Instead, the plant and corporate levels have often been narrowly treated. Strategic instruments created for the determination of the business strategies could also be used when determining the corporate or plant asset strategy. Of course, some further development has been required. In addition to the above mentioned issues there is a lack of economic and technical models in order to improve decision making concerning expansion, modernisation and replacement investments.

In the capital-intensive industry, considerations dealing with investment policies and maintenance development are a part of the strategic and operational decision making. In the rapidly changing business environment and increasing profitability demands, the industry would profit from new business approaches like 'broader scope asset management'.

6 FUTURE RESEARCH

During the roadmap process carried out in 2004, we identified several important topics for the future research efforts. Some examples have been presented below.

6.1 Capacity management

There is a generic need to construct a calculation technology and software solution for the optimisation of a company’s or plant’s capacity and use of that capacity. This solution should take into account uncertainty in the middle-term planning horizon and changing business environment.

6.2 Improvement investments and maintenance

A dynamic and probabilistic model to support the original LCP-objectives is required. The options available are development investments in existing production units and maintenance actions. These models should take into account various business constraints and changes in them, such as the economic age of the equipment, environmental risks, safety, price trends, technical development etc.

6.3 Risk management for Asset management

This research theme covers an overall risk management system for the physical assets, taking into account the business objectives, changes in the business environment, viewpoints of various joint-parties, the balanced governance of potential opportunities and versatile risks. The resulting risk management instruments should make it possible to carry out qualitative and quantitative risk analyses for the production system.

6.4 Intangible assets

This research area covers ICT-systems, expert knowledge, tacit knowledge and human interaction required to take the full use of the physical assets.

7 REFERENCES

[1] Siitonen S. (2003) Impact of Globalisation and Regionalisation Strategies on the Performance of the World's Pulp and Paper Companies. Doctoral Dissertation. Helsinki School of Economics. Helsinki.

[2] Komonen K., Kortelainen H., Laakso K., Rosqvist T., Räikkönen M. (2005) Physical Asset Management, a Challenge for Finnish Capital-Intensive Industry. in Industrial Systems Review. VTT Industrial Systems. Technical Research Centre of Finland. Espoo. Finland, pp. 78-84.

[3]

[4] Mitchell J. S. (2002) Physical Asset Management Handbook. 3. Edition. Clarion Technical Publishers. 291 p

[5] Peltonen M., Kortelainen H., Kuukkanen R. and Virtanen I. (2002) A special tool for life cycle cost and profit analysis can be an invaluable support for investment decisions, Pulp and Paper Europe. Paperloop.com

[6] Komonen, K. (2002) A cost model of industrial maintenance for profitability analysis and benchmarking. International Journal of Production Economics, Vol 79, pp.15-31

[7] Harrigan K. R. (2003) Vertical Integration, Outsourcing and Corporate Strategy. Washington. BeardBooks p. 372. ISBN 1-58798-190-4

[8] Komonen K. (2006) Availability Guarantees and Physical Assets Management: Empirical Evidence of the Impact of Some Underlying Factors. *RAMS 2006. Conference Proceedings. Newport Beach. CA, USA*

[9] Komonen, K. (2002) Views on performance indicators and benchmarking in industrial maintenance. Maintenance 5, pp. 52-56

[10] Kelly, A. (1998) Maintenance organization and systems. Business –centred maintenance. Great Britain. Butterworth-Heinemann. 280 p. ISBN 0 7506 3603 3

[11] Komonen K. (2004). Technical and economic assessment, development and optimising techniques of industrial maintenance, in Holmberg, K.; Komonen, K.; Oedewald, P.; Peltonen, M.; Reiman, T.; Rouhiainen, V.; Tervo, J.; Heino, P. 2004. Safety and Reliability - Technology Review. VTT Industrial Systems, Espoo. 80 s. + liitt. 4 s. Research Report BTUO43-031209.

[12] Nakajima S. (1989) TPM Development Program: Implementing Total Productivity Maintenance. Portland, Oregon. Productivity Press, Inc. 403 s. ISBN 0-195299-37-2

[13] Scarf, P.A. (1997) On the application of mathematical models in maintenance. European Journal of Operational Research 99, pp. 493-506

[14] Blischke, W., R. & Murthy, D., N., P. (1994) Warranty Cost Analysis. New York. Marcel Dekker, Inc. 731 p. ISBN 0-8247-8911-3

[15] Blischke, W., R. & Murthy, D., N., P. (1996) Product Warranty Handbook. New York. Marcel Dekker, Inc. 925 p. ISBN 0-8247-8955-5

[16] Swanson L. (1997) An empirical study of the relationship between production technology and maintenance management. International Journal of Production Economics. Vol 53, pp191-207

[17] Swanson L. (2001). Linking maintenance strategies to performance. International Journal of Production Economics, Vol 70, pp. 237-244

[18] Bhat, V.N. (2000) The determinants of maintenance expenditures in chemical companies. Journal of Quality in Maintenance, Vol 6, No. 2, pp. 106-112

[19] Komonen, K. (1998) The Structure and Effectiveness of Industrial Maintenance. Ph.D. Dissertation. Helsinki University of Technology, Department of Industrial Economics, Espoo. in Finnish

E-FAMEMAIN: A BENCHMARKING WEB-TOOL FOR O&M MANAGEMENT

Kari Komonen

VTT (Technical Research Centre of Finland), P.O. Box 1000, FIN-02044 VTT, Finland.

Abstract: E-Famemain is a new kind of a tool for benchmarking, which is based on many years' research efforts. It helps to evaluate users' performance in operations and maintenance by making industrial plants comparable with the aid of statistical methods. The system is updated continually and automatically. It carries out automatically multivariate statistical analysis when data is entered into the system, in addition to many other statistical operations. The system consists of four sections: data input section, positioning section, locating differences section and best practices and planning section.

Key Words: Maintenance, Benchmarking, Overall Equipment Effectiveness, Availability, Maintenance Costs,

1 INTRODUCTION

The success of the maintenance function can be evaluated (simplifying) with the aid three factors: the product of maintenance efforts, cost-effectiveness of maintenance function and the quality of maintenance processes. The product of the maintenance activities can be measured e.g. with the aid of overall equipment effectiveness (OEE), availability or reliability of the machinery. Typical measures for the cost-effectiveness of the maintenance function are 'maintenance costs in relation to plant replacement value' or 'maintenance costs in relation to production volume'. The quality of the maintenance processes can be evaluated with such indicators as planning rate, customer satisfaction, employee satisfaction, accident rate or hazard rate etc. The significance of machinery's availability performance and overall cost-effectiveness has increased because of the harder competition in many industrial branches. This trend has caused a demand for a higher OEE- level and higher efficiency requirements.

According to the management of many industrial enterprises, 'benchmarking' is a powerful tool for the development of the plant operations. However, among many problems concerning benchmarking one challenge is to make companies or plants comparable. This challenge is sometimes used as an excuse to explain away the worse ranking in a comparison. In this paper, we describe the system of benchmarking, which aims to make various production units comparable with the aid of statistical methods based on many years' theoretical and empirical research. The positioning of the production systems is one of the instruments introduced in this paper, but the other tools lean also on the results of the positioning tool.

Obviously the business environment, cost structure and technology of any production system may have a remarkable impact on the content and level of availability requirements. In the capital-intensive industry usually the high utilisation (operating) rate of the production equipment is needed for profitable operations. Also, low profit margins put pressure on increasing the production volume, which again increases the need for the higher utilisation rate and availability. The strategy of the company may also have a significant impact on availability requirements (Komonen 2006) [1].

Costs to maintain a required availability level vary also depending on the industrial sector and specific features of the plants in question. It is often beyond the scope of the production and maintenance management's determining power to influence in the short run, for example, the lack of scale economies. The same applies to the operating rate of the production equipment (e.g. Komonen 1998, 2002a, 2002b, 2006) [2,3,4,1]. Similar empirical studies are not too rich. However, there are some interesting papers concerning the business and technological determinants of maintenance expenditures e.g. Swanson 1997 and 2001 [5,6] and Bhat 2000 [7].

When determining the required input and output, there is always a trade off situation with maintenance efforts and OEE or availability. Therefore it would be very important to know the total costs of availability performance or to have the tool to make compound analyses. In certain circumstances it may be more profitable to decrease maintenance costs than to improve OEE level.

2 SOME VIEWS ON THE USE OF PERFORMANCE INDICATORS

One of the important tools for asset management is the system of the performance indicators. It is needed at all the levels of asset management. Often the systems of the key indicators are lists, which are grouped according to subject matters but not in any way linked to each other. Surely, there are classifications which aim to organise the performance indicators into various classes such as economic, organisational and technical. The aim of this chapter is to introduce the system of maintenance key indicators, which supports the building of a benchmarking system. This system helps a user to grasp the purpose and significance of various key figures.

The simplified system of the performance indicators can be classified in a hierarchical manner. The following structure is the modified version of those presented earlier by Komonen (e.g. 2002b) [4].

- business-oriented objectives
    1. first level objectives of production
    2. second level objectives of production
    3. objectives for maintenance
    4. exogenous factors for production and maintenance (external factors)
    5. intermediate internal objectives for maintenance (follow-up variables)
    6. action variables of maintenance function (means for improvement)
    7. internal descriptive (explanatory) variables

Such **fundamental business goals** as ROI, life cycle profits or profit margin belong to the top level of hierarchy. The following level of objective variables contain **the first level objectives of the production function** as 'total lost production opportunities', which stands for lost production in money terms due to external disabled time and downtime due to production and maintenance down time. Other objectives at this level may be 'turnover of capital' or 'availability performance costs' (unavailability costs plus maintenance costs). The following level of the performance indicator system is **the second level objectives of the production system**. This category has been split into three objective areas: overall equipment efficiency (OEE), the costs of the production system and the quality of the production processes. **The maintenance objectives** are metrics for the performance of maintenance operations: e.g. availability of machinery due to maintenance, sum of unavailability costs and maintenance costs ("dependability costs") or maintenance costs as a % of the estimated plant replacement value (PRV). **Exogenous independent factors** are indicators, which help management to evaluate the state of technology or business environment and at the same time are beyond the scope of production or maintenance managers to determine, but which have a considerable impact on the other objectives. Examples of this kind of factors are the utilisation rate of the production machinery, production volume, the size of the production equipment, the integration level of the production process etc. **Intermediate internal objectives** are performance indicators which should be measured and followed up since they may give more information about the development needs and may act as intermediate objective variables in order to reach the main objective.

Table 1. The O&M focused system of the plant level asset management indicators

Fundamental business objectives	*First level objectives of production*	*Second level objectives of production*	*Maintenance objectives*	*Exogenous Factors*	*Intermediate objectives*	*Means, action variables*	*Descriptive variables (internal explanatory)*
	Total lost production opportunities (€)	*OEE*	e.g. * availability * performance rate * quality rate	e.g. *operating rate * technology * integration level * capital intensity	e.g. * MTTF * MWT * MTTR * MTTM	e.g. * preventive maint. * planned maint. * spare parts * contracting rate	e.g. * inventory turnover * proportion of mehanical maintenance
**ROI, BSC,* **Growth,* * *Market share etc.*	*Production dependability costs (unavailability costs+upkeep costs)*	*Production costs*	e.g. * maintenance costs / PRV * maintenance costs / output	e.g. * operating rate * technology * scale * integration rate * capital intensity	e.g. * MRT * MTTR * MTTM * flexibility * spares / PRV	e.g. * preventive maint. * planned maint. * contracting rate	* material costs * allocated hours * unit labor costs * maintenance costs/ plant turnover
	Turnover of capital	*Quality of production processes*	e.g. * internal customer satisfaction * planning rate * job satisfaction	e.g. * technology * structure of production system	e.g. * MWT * MTTM * Accidents * Absenteeism * Claims	e.g. * feedback to clients * multi-skilled workers * keeping promises * improvement maintenance	e.g. * age structure of personnel

For example, the mean time to restoration (MTTR) is an intermediate objective when the main objective is to minimise downtime costs. These variables are objectives, because a production or maintenance manager cannot influence them directly. **The action variables of the maintenance function** are tools of the production or maintenance managers, with the aid of which objectives are achieved. Preventive and improvement maintenance, outsourcing, operator maintenance etc. are good examples of these tools. **Internal descriptive (explanatory) variables** give additional information, for example, about the cost and organisation structure, cost level and capital intensity of the maintenance function.

Although the above mentioned performance indicators cover partly both the production and maintenance functions, the main focus is on maintenance activities. The system of the key figures could be illustrated by several centric zones (Komonen 2002a) [3], but here a matrix presentation is applied. The results presented in Table 1 are examples. The table does not include all the possible indicators. Owing to lack of space some variables are in the form of a ratio and some in the form of a concept. However, for all the variables presented it is possible to construct an indicator in the form of a ratio or a value.

2.1 Exogenous factors a priori

The top management of the companies should be interested in the structural causes (exogenous factors) of the maintenance product and cost level in order to develop the operations of the company. The maintenance management should be interested in the impact of the exogenous factors in order to draw correct conclusions concerning the maintenance function when using benchmarking.

In the studies mentioned before, multiple regression analyses were used to estimate the causalities between the exogenous factors and the key performance indicators (objectives). Logarithmic forms of the variables were often needed in the analyses, because the model to be estimated was non-linear. Logarithmic transformations have also other statistical and mathematical advantages (e.g. estimated parameters measure the elasticity of the variables). The explanatory power of the models (adjusted R^2) was high (mainly between 40% and 90 %) and the estimated parameters were statistically highly significant (e.g Komonen 1998, 2002a, 2006) [2,3,1].

As mentioned above, different kinds of production units or plants can be made comparable with the aid of exogenous factors. Thus, for example, objectives such as 'overall equipment effectiveness' (OEE) or availability are dependent on the external factors, which cannot be influenced in a short time span. Examples of the exogenous factors can be a priori the following variables (the signs + and – indicate the direction of the impact):

- integration level of the production system (+)
- operating rate and /or shift work rate (+)
- technical capital intensity (plant turnover / plant replacement value) (+)
- redundancy (+)
- severity of the production environment (+)

In the case of maintenance costs in relation to the equipment replacement value, exogenous factors may be the following:

- integration level of the production system (which has, in practice, often parallel impacts to maintenance costs as replacement value) (-)
- scale (replacement value of the production equipment) (-)
- production volume / production equipment (replacement value) (+)
- shift work rate, operating rate (+)
- industry –dummy variables (industry specific factors) (-, +)

In the case of 'maintenance costs / production volume' the results were similar, but the direction of impact was partly different:

- integration level of the production system (-)
- scale (replacement value of the production equipment) (-)
- production volume / production equipment (replacement value) (-)
- production volume (-)
- industry –dummy variables (industry specific factors) (-, +)

The ratio 'maintenance costs / plant replacement value' tells us how much it costs to maintain a certain type and size of equipment or facility. When using this ratio we should make a difference between the total costs of maintenance and the maintenance costs of the machinery. In certain branches the proportion of real estate of the total fixed capital may vary a lot, and therefore this index may give misleading results, at least, as far as the maintenance costs of the machinery are concerned. Altogether, the problem with the above ratio is the determination of the replacement value. Very often the fire insurance value is a good estimate of replacement value. The advantage of this ratio is that it makes it possible to compare various kinds of plants within or between industrial branches. The disadvantage of this ratio is that it does not take into account the availability of the production equipment

The ratio 'maintenance costs / production volume' takes implicitly into account production losses and both the factors can be defined quite exactly. The disadvantage of this ratio, however, is that very often the output of the production is very difficult to determine physically. Thus, in many branches benchmarking with the aid of this index is difficult. In some branches it may be possible to use substitutes, for example the consumption of energy or raw material. Another challenge is that the production volume is often more flexible than the maintenance costs.

3 EMPIRICAL EVIDENCE CONCERNING EXOGENOUS FACTORS

There exists empirical evidence of the impact of the business environment and technological determinants e.g. on OEE (overall equipment effectiveness), availability and maintenance costs (e.g. Komonen 1998, 2002b, 2006) [2,4,1]. The structure of the production system has an impact on availability requirements as stated above. If a machine is the bottleneck of the production system, it is obvious that availability objectives are higher than in the case of a machine, which is not in the main production line and which has a low operating rate (Komonen 2002) [8]. "An indicator to model almost any production system is the degree of integration (plant replacement value divided by the number of front line production operatives in day shift), which is a multipurpose meter to measure mechanisation rate, continuity of processes and scale economies. Generally speaking, we can state that e.g. in the chemical industry and in the pulp and paper industry the integration level is high and in the manufacture of metal products it is lower. The impact of the degree of integration on the above mentioned performance indicators is demonstrated in Table 2 and is illustrated in Figure 1. The higher the degree of integration of the production systems the higher availability performance they show. The same applies to the planning rate of the maintenance activities" (Komonen 2006) [1]. In Figure 1 the predicted and observed values equal on the diagonal. Cases above the diagonal represent better performance than expected.

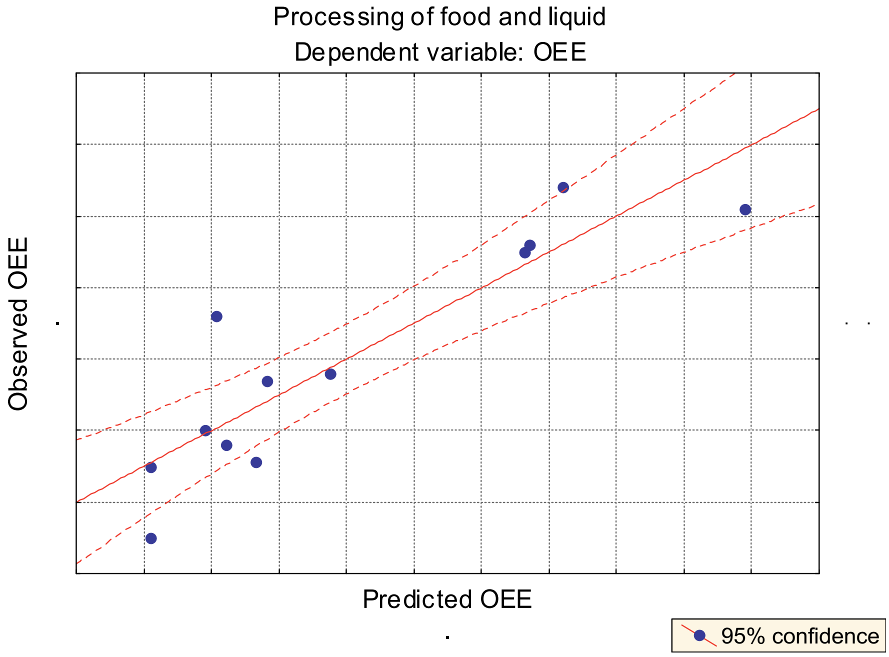


Figure 1. The predicting power of the integration level on OEE in the processing of food (Komonen 2006) [1]

Table 2

The variables having an impact on OEE in the processing of food and liquid (Komonen 2006) [1]

Processing of food and liquid (N=12)		
Dependent variable: OEE		
Adjusted R^2=0,721	$F(3,8)$=10,498	p<,0037
	t(8)	p-level
Intercept	6,608	0,0002
Degree of integration	4,265	0,0027
Production equipment replacement value	-2,183	0,061
Plant turnover / plant replacement value	-1,328	0,221

The relationship between the relative maintenance costs and some exogenous factors concerning the same industry as above is demonstrated in Figure 2. In Figure 2 predicted and observed values equal on the diagonal. Cases below the diagonal represent better performance than expected. The specific impact of various exogenous variables is shown in Table 3.

Table 3

The variables having an impact on the relative maintenance costs in the processing of food and liquid (Komonen 2006) [1].

Processing of food and liquid (N=34)		
Dependent variable: Production equipment, Maintenance costs / Replacement value		
Adjusted R^2= ,542	F(4,29)=10,773	p<,00002
	t(29)	p-level
Intercept	3,865	0,0006
Replacement value	-4,386	0,00014
Degree of integration	-2,275	0,030
Plant turnover / plant replacement value	1,859	0,073

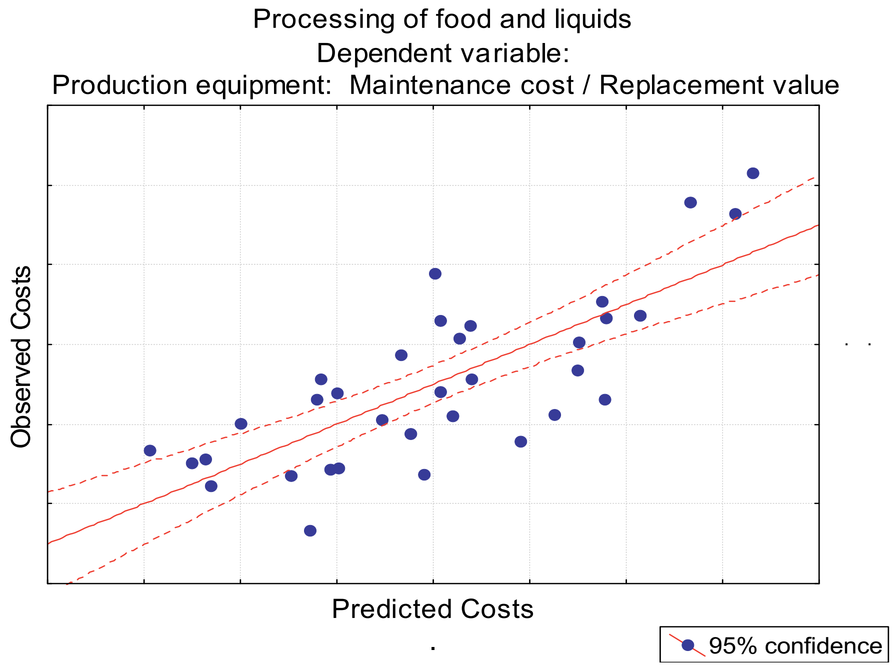


Figure 2. The predicting power of exogenous variables on the relative maintenance costs in the processing of food and liquid (Komonen 2006) [1].

It was interesting to notice that in the processing of food and liquid, the variables explaining maintenance costs were parallel to the variables explaining variation in OEE, although the direction of the influence was not always the same. However, in the first case the replacement value was the most significant indicator.

The results presented above offer an opportunity to judge whether a plant in question has been successful or less successful. This evaluation can be carried out with the aid of a two by two table illustrated in Figure 3. In Figure 3 the measure for production effectiveness is availability instead of OEE. The deviation of availability from the diagonal has been presented on the vertical axis and the deviation of maintenance costs from the diagonal on the horizontal axis. In the case of the upper right-hand corner the plant in question is obviously successful, and in the opposite corner unsuccessful. In the case of the upper left-hand corner, a good strategy may be to concentrate on maintenance costs keeping availability at the present level. The lower right-hand corner represents a situation where a good option would be to focus on the availability of the production equipment.

The ultimate goal surely is that managers would be able to calculate the dependability (availability performance) costs of plants. Therefore, at least, unavailability costs, maintenance costs and investments on availability performance should be included. However, in practice, usually maintenance costs and availability as a time concept (%) has been followed. When optimising maintenance operations there is always a trade off situation with maintenance efforts and availability. Therefore it would be very important to know the total costs of availability performance.

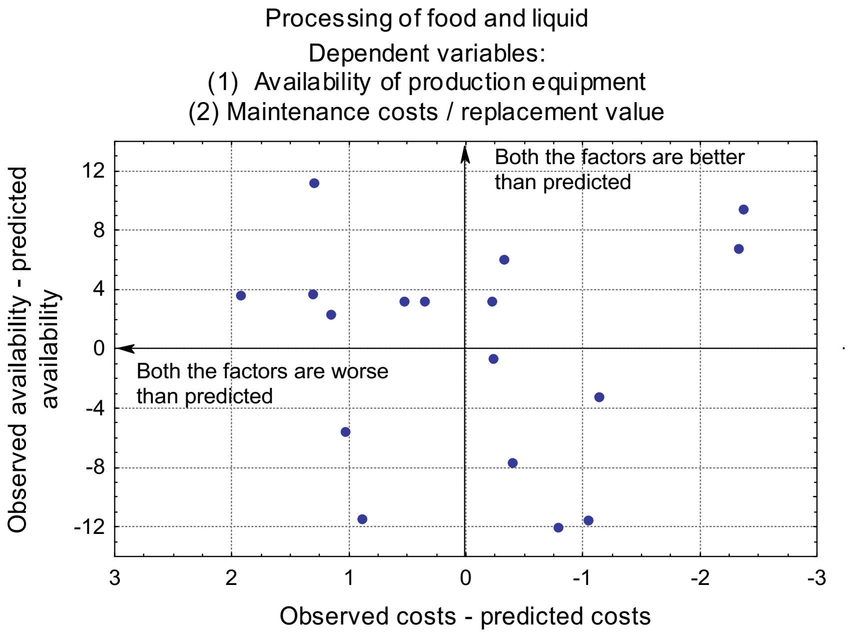


Figure 4. Positioning of availability and maintenance costs (Komonen 2006) [1]

Total costs are, however, very seldom calculated. As an example, the correlation between the total availability performance costs and utilisation rate of the production equipment has been illustrated in Figure 5 (Komonen 2002) [4]. The causality below is empirical and it has been calculated for only one industrial branch. Therefore, the mechanism behind this phenomenon is not clear. However, it can be stated that availability should be high when operating rate is high.

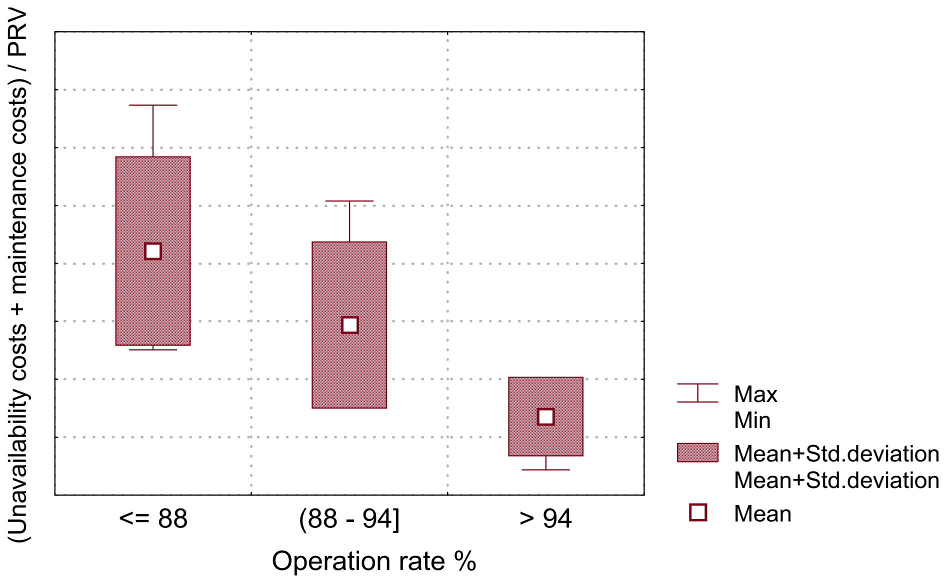


Figure 5.Relationship between the total availability performance costs and the utilisation rate of the production equipment.(PRV=Plant Replacement Value)

On the basis of the empirical examination and on the basis of estimated equations, a positioning tool can be produced for various industrial sectors. The commonly applied benchmarking procedure has been the comparison of the average values of the particular industrial sector with the company's own values. However, in the field of industrial maintenance this kind of benchmarking does not help very much. The "owners" of the best figures are too "proud" of their performance and the "owners" of the poor ones are on the brink of despair. According to the results of this research, the above-mentioned average value based methods are inappropriate. Even the misuse of benchmarking values and opportunistic behaviour are possible.

4 WEB TOOL FOR BENCHMARKING

4.1 Use of positioning results

With the aid of positioning results the sample of plants in the benchmarking data-base can be split down into two groups: successful plants and unsuccessful plants. These two groups can be used when trying to identify causes for successful or less successful performance of the individual plant.

Applying positioning to all the cases in the sample it is possible to compare industry-wise the plant's mode of operations with the average plant or with the successful plants or with the less successful plants. Comparison can be made with the aid of several economic, technical and organisational indicators. Examples of the modes of operations are the amount of operator maintenance, preventive maintenance, contracting and improvements or the use of criticality analyses. By following the techniques introduced above, it is possible to find some hints of which kind of modes of operations lead to successful results. Therefore the proper way to improvements depends on the plant's present results and present mode of operations.

As soon as the differences in the modes of operations have been identified it is possible to investigate which modes of operations are effective and which are not. The benchmarking tool introduced below is based on the above mentioned research results concerning both positioning and successful modes of operations.

4.2 General description of benchmarking system

The benchmarking system consists of four sections or tools. (1) 'Data input section' is an electronic form to insert the data of the plant in the database. An incorporated test tool will give the user a notice in the case of the most probable mistakes. (2) In the 'positioning section' there is an option to evaluate the success of the plant's maintenance function branch-wise (by industrial sector) with the aid of several indicators such as OEE, availability of machinery, production losses due to maintenance, maintenance costs as a percentage of estimated plant replacement value and many other. In the 'locating differences section' it is possible to compare the plant's mode of operations with the successful plants, with the less successful plants and with an average plant in the industrial sector in question. Comparison can be made with the aid of several economic, technical and organisational indicators. In the 'best practices and planning section' the user of the benchmarking system may find more hints of which kind of the mode of operations leads to successful results. The structure of the tool is illustrated in Figure 6.

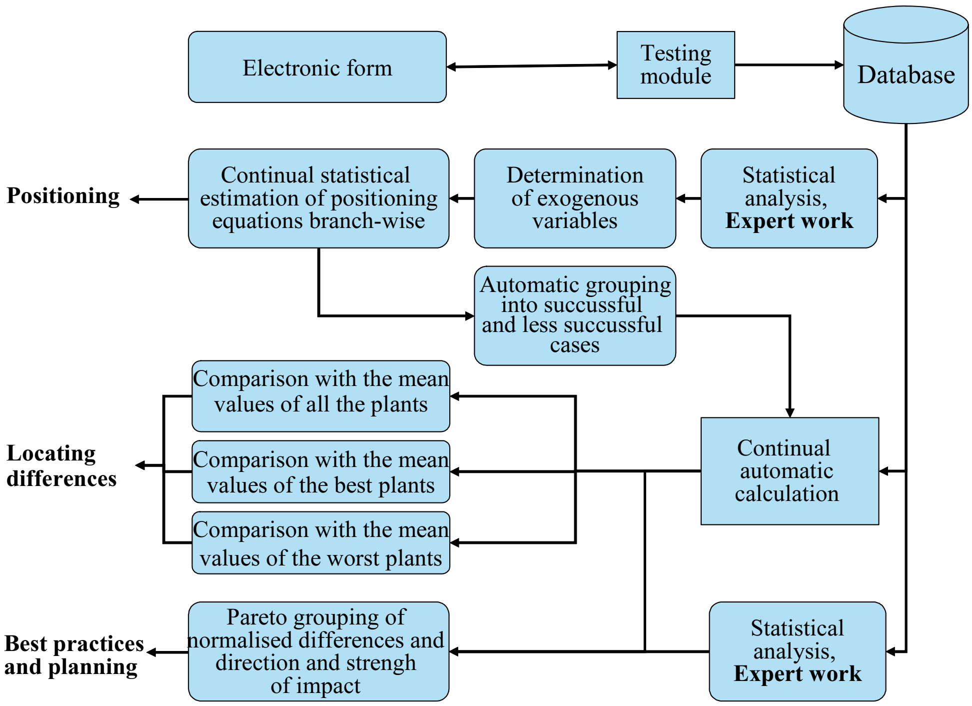


Figure 6. Schematic description of the benchmarking system

The system is automatic and runs all the automatic calculations recieving commands from the username and password entered in the system. All the needed calculations are carried out when changes occur or the user commands to execute any task. Simulations are also possible options, for example, for planning purposes. The determination of the variables, which should be taken into account when the system executes automatic multivariate analyses for the positioning, is an expert's job. The same applies to the determination of the impact (direction and strength) of the modes of operations on the selected objectives. The users of the system are able to see only their own data and statistical key figures concerning the industrial branch in question.

4.3 Detailed description of various tools of the benchmarking system

The web-tool for benchmarking consists of four sections.

"Data input section" serves the user with an electronic form to insert data concerning the plant's maintenance function. An incorporated test tool will give the user a notice in the case of the most probable mistakes. Data entered by the user is a precondition for the broader use of this benchmarking system. Data input sheet is illustrated in Figure 7.

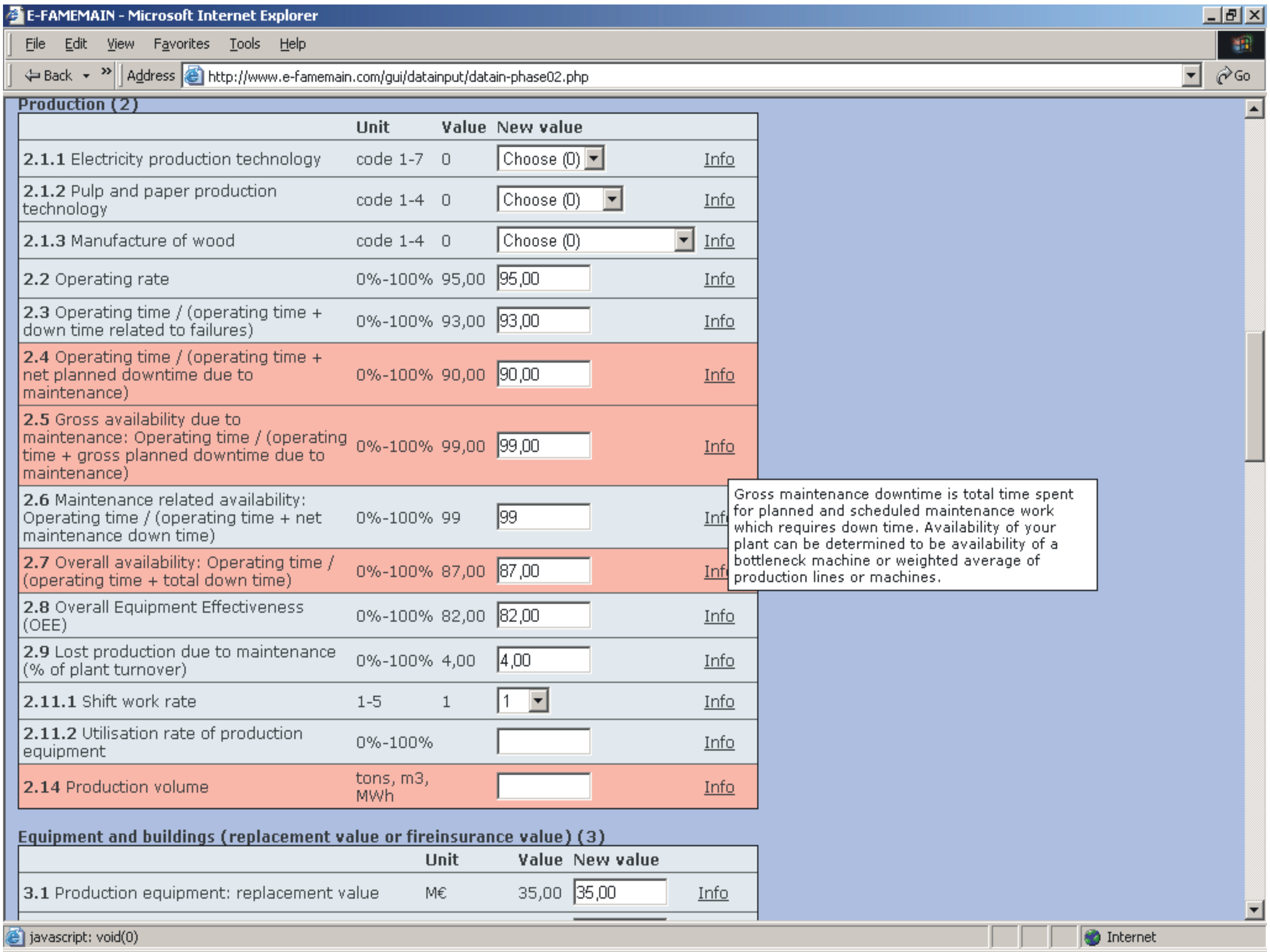


Figure 7. The illustration of the data input page

In the **"positioning section"** the users are able to evaluate the success of their maintenance function branch-wise (by industrial sector) with the aid of several indicators such as availability of machinery, production losses due to maintenance, maintenance costs as a % of the estimated plant replacement value and many other. The above mentioned indicators are affected by several exogenous factors which are beyond the scope of the plant management or maintenance managers to determine, but still they have a considerable impact on the performance indicators. Examples of exogenous factors are process severity, plant size, utilisation rate, age of the plant, technology. These factors are taken into account when evaluating one's success with the aid of multivariate statistical models. The conclusion page of the positioning section is illustrated in Figure 8.

As stated earlier in the positioning section, various plants are made comparable with the aid of exogenous factors. The benchmarking tool will take care of this analysis automatically for the user. The developed software calls plants' data from the data base, takes into account influencing factors and calculates for the users an expected, predicted value for each performance indicator the users have chosen. Users can also make simulations, how things would change if they were able to change factors which are or are not under their own control. This benchmarking software also groups all plants into successful and less successful subgroups.

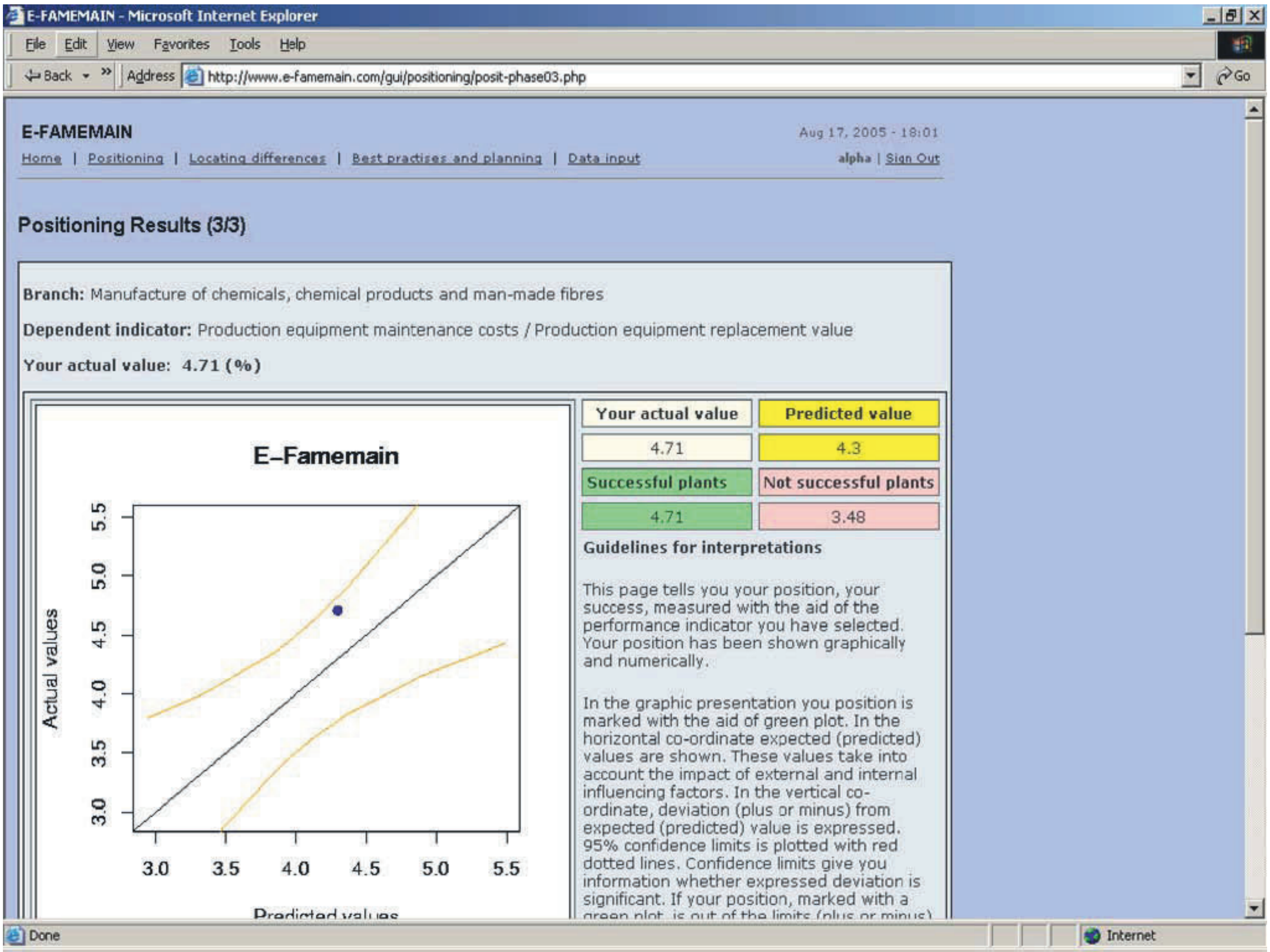


Figure 8. The conclusion page of positioning

The above page tells the users their position, their success, measured with the aid of the performance indicator they have selected. Their positions are shown graphically and numerically. In the graphic presentation their positions are shown as a green plot. The horizontal co-ordinate shows predicted values. These values take into account the impact of the external and internal influencing factors (exogenous factors). The vertical co-ordinate shows the actual values of users. 95% confidence limits are plotted as red lines. The confidence limits indicate whether the expressed deviation is significant. The position of the green plot marking plant's performance can be changed by using the simulating tool in the previous page.

In the **"locating differences section"** the user is able to compare the plant's mode of operations with the successful plants, with the less successful plants and with an average plant in the industrial sector in question. Comparison can be made with the aid of several economic, technical and organisational indicators. The conclusion page of "locating differences section" is illustrated in Figure 9.

The objective of this analysis is to find probable causes for deviations from the expected (predicted) values in the positioning section. The benchmarking tool will take care of this analysis automatically for the user. The software calls all the needed data from the data base and carries out positioning for each case and in this way groups all the cases into successful and less successful subgroups. Finally, the tool calculates means and deviations for the groups of all cases, successful plants and less successful plants and for each performance indicator the user has chosen. Plants' own actual values for the selected indicator is shown as a green plot. The user can also make simulations about how things would change if he would able to change factors which are or are not under his own control.

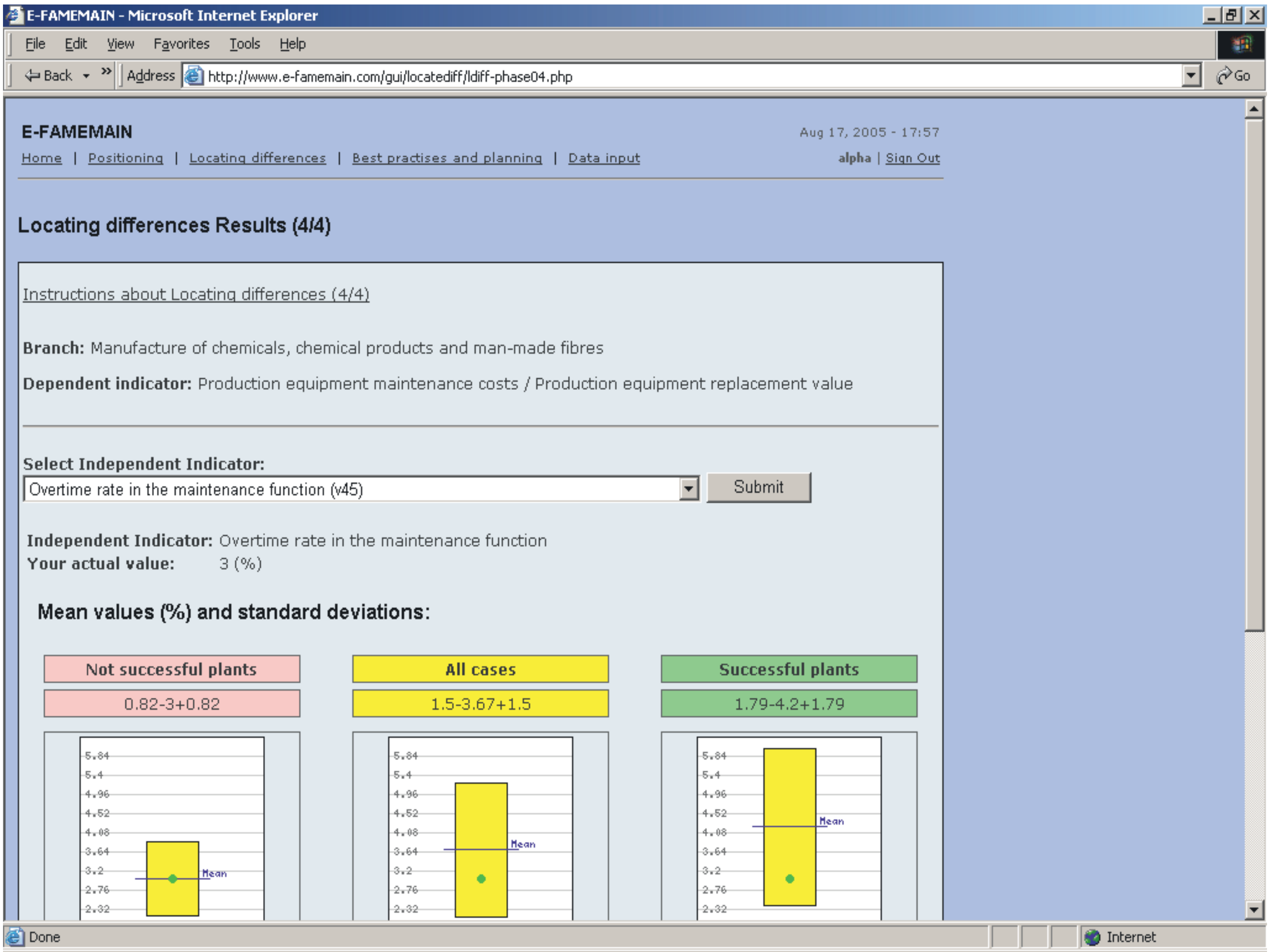


Figure 9. The conclusion page of locating differences

In 'best practices and planning section' the user can find some hints of which kind of a mode of operations leads to successful results. Since the proper way to improvement depends on the user's present results and present mode of operations, this section includes also a management tool to help in the planning process.

The user is able to compare the plant's mode of operations with the successful plants. The comparison is a summary of the 10 largest deviations in the form of Pareto-diagram and the comparison can be carried using several economic, technical and organisational indicators. The objective of this analysis is to find a course of actions which should be carried out in order to improve the performance of the objective variables (e.g. OEE, availability or cost effectiveness). Deviations can have positive or negative values.

Since the statistical performance of the successful companies is based on the results of positioning, those results are presented here once again in order to make the interpretation easier. Thus, the simulation tool is also available in this section. The differences between the mean values of the most successful plants and the users' own indicator values have been normalised. The graphical conclusions are presented in Figure 10. The effect of the indicators in question on the dependent variables (e.g. OEE, availability or cost effectiveness) is indicated in the same table (direction and strength). The causalities between independent and dependent variables are studied using the sample entered into the data base. Clear causalities can not always be identified.

The strength of impact is presented with the aid of five levels: no impact, small influence, medium influence, significant influence and very significant influence. The direction of the effect denotes what happens to an objective value (e.g. availability) if the numerical value of an independent indicator (indicator which depicts the mode of operations) increases.

In the following page the summary of the largest deviations and their influence on the performance indicators are presented in a two by two table. The greatest attention should be focused on the factors which indicate the largest deviations and strongest impact. Further, less attention should be focused on the factors having smaller deviations and smaller impact. If e.g. the availability and maintenance costs are studied simultaneously, the actions needed may be conflicting. In this case, focus should be directed to those activities which have a positive impact on both the factors.

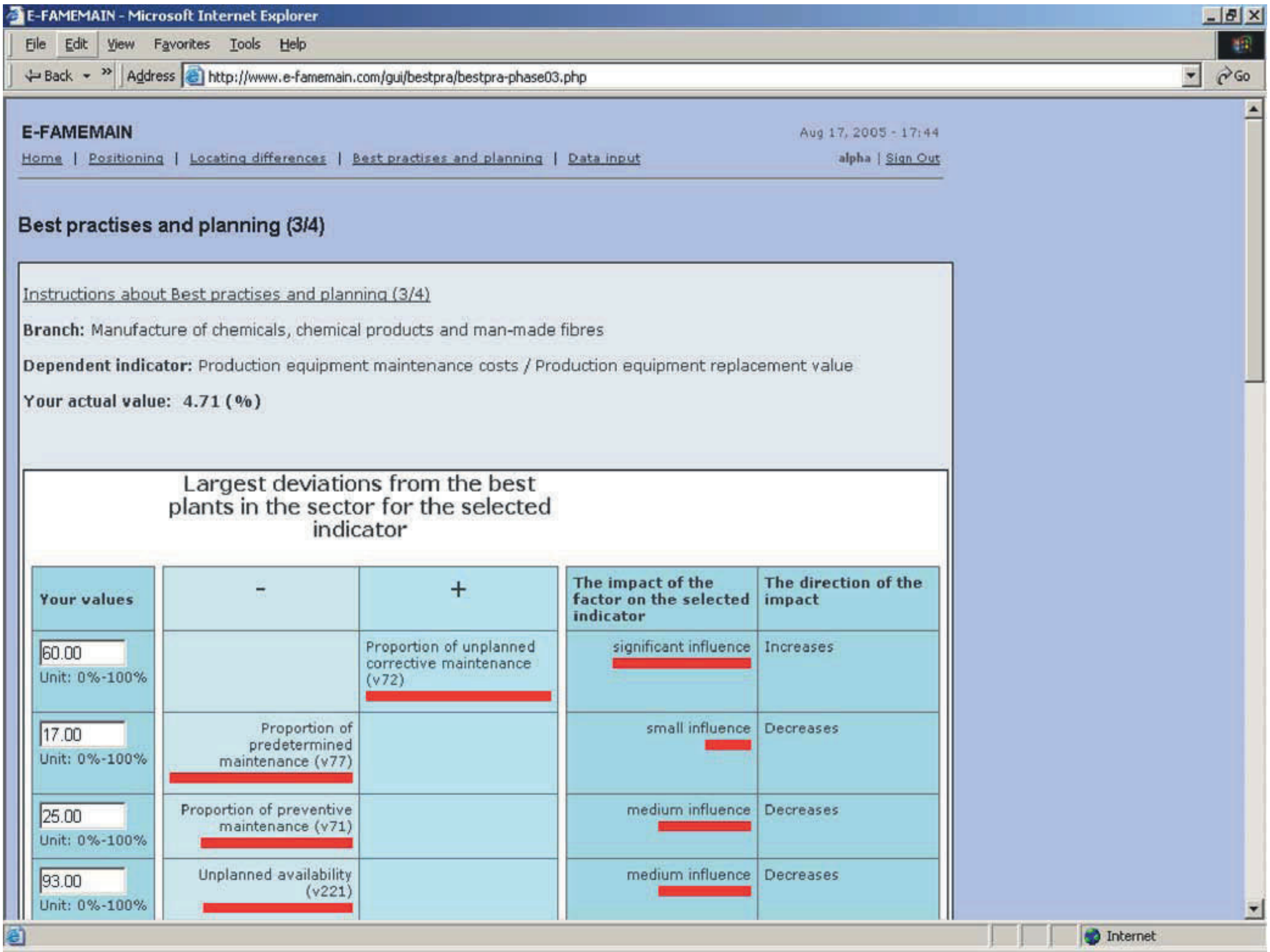


Figure 10. The first conclusion page of the best practices and planning section

5 DISCUSSION

The introduced web benchmarking tool is based on the research concerning causalities between objective variables and exogenous influencing factors, and on the research concerning causalities between the successful plants (determined with the aid of the positioning) and modes of operations determined by action variables such as preventive maintenance or training of personnel. The results of the first mentioned causalities have been clear and the samples have been sufficiently large. However, the latter causalities are not as clear. The main reason here is the strong data erosion due to diverse answering activity to various questions, that is, various indicators.

Regardless of the above mentioned problems some clear results have been achieved. In several industrial branches, the means which improve the product of maintenance activities such as OEE or availability increase the relative costs of the maintenance function and vice versa. Sometimes the results are opposite to expectations. Therefore, the optimisation of the efforts is crucial. The quality of data and sufficient data are also crucial issues when improving benchmarking tools. In empirical research there is always a trade off between the detailed data and the number of cases in a sample. The ability of the companies to deliver data for a large number of sophisticated indicators is limited. The introduced benchmarking system is quantitative in nature. There is surely a need to add qualitative features to the system. However, as soon as the scope of the tool is broadened to the qualitative field there is a need to change the research strategies, and that means challenges in the architecture of the benchmarking tool and also problems in willingness to deliver data.

For the improving the benchmarking system, the use of new performance indicators should be activated. These indicators should facilitate and intensify co-operation between maintenance and operator functions. In order to draw correct conclusions when using those indicators, more research concerning causal relationships between objectives and exogenous factors (and also action variables) should be carried out. These causalities should be used when constructing benchmarking systems and tools. There is also a development process going on to add some new technical features to the system.

6 CONCLUSIONS

The significance of machinery's availability performance and overall cost-effectiveness has increased because of the harder competition in many industrial branches. This trend has caused a demand for a higher OEE- level and higher efficiency requirements. Obviously the business environment, cost structure and technology of any production system may have a remarkable impact on the content and level of availability requirements.

According to the management of many industrial enterprises, 'benchmarking' is a powerful tool for the development of the plant operations. However, among many problems concerning benchmarking one challenge is to make companies or plants comparable. This challenge is sometimes used as an excuse to explain away the worse ranking in a comparison. In this paper, the benchmarking system to avoid this problem has been introduced. The positioning of the production unit seems to work well and the quality of the positioning system has improved during the years of research.

The identification of the best practices has been a harder task because of the strong data erosion caused by inadequate filling of the questionnaires. There is also a trade off situation with maintenance efforts and OEE or availability, which makes it more difficult to draw clear conclusions concerning the effective modes of operations. Therefore, there is still room to make better compound analyses.

The introduced benchmarking tool is the result of more than ten years' systematic research. It leans on the following phases.

(Phase 1) Several years research work on the causalities between various performance indicatos based on the annual data collection in Finland.

(Phase 2) Several years development and standardisation of performance indicators to build up the common system of indicators for benchmarking activities in Europe. The hierarchy of indicators used in the e-Famemain benchmarking system has been designed together with Finnish maintenance society (KPY) and Swedish Maintenance Society (UTEK).

(Phase 3) The development of the first generation bechmarking web tool, which was installed in the year 2000.

(Phase 4) Identification and description of the best practices: causalities between various practices and objective variables, and the development of methods to transfer the research results into the properties of the benchmarking system

(Phase 5) Specification and programming of the web-tool during the years 2004 and 2005.

7 REFERENCES

[1] Komonen K. (2006) Availability Guarantees and Physical Assets Management: Empirical Evidence of the Impact of Some Underlying Factors. *RAMS 2006. Conference Proceedings. Newport Beach. CA, USA*

[2] Komonen, K. (1998) The Structure and Effectiveness of Industrial Maintenance. Ph.D. Dissertation. Helsinki University of Technology, Department of Industrial Economics, Espoo. in Finnish

[3] Komonen, K. (2002a) A cost model of industrial maintenance for profitability analysis and benchmarking. International Journal of Production Economics, Vol 79, pp.15-31

[4] Komonen, K. (2002b) Views on performance indicators and benchmarking in industrial maintenance. Maintenance 5, pp. 52-56

[5] Swanson L. (1997) An empirical study of the relationship between production technology and maintenance management. International Journal of Production Economics. Vol 53, pp191-207

[6] Swanson L. (2001). Linking maintenance strategies to performance. International Journal of Production Economics, Vol 70, pp. 237-244

[7] Bhat, V.N. (2000) The determinants of maintenance expenditures in chemical companies. Journal of Quality in Maintenance, Vol 6, No. 2, pp. 106-112

[8] Komonen K. (2002) "Long-term availability guarantees and capital asset management ", *Euromaintenance 2002, Conference proceedings,* Rajamäki, Finnish Maintenance Society, 2002

A METHOD FOR MANAGING THE AVAILABILITY IMPROVEMENT EFFORTS

Barabady J [a] and Kumar U [b]

[a, b] *Division of Operation and Maintenance Engineering, Luleå University of Technology.*

Abstract: Availability is an important characteristic of a repairable system. When the availability of system is low, efforts are needed to improve it. Any improvement in the availability of a system is associated with the requirement of additional effort and cost. Therefore, it is essential to use methods or techniques for availability allocation amongst various components/subsystems of a system with the minimum effort and cost.The concept of importance measures could be used to prioritize the components or subsystems for the availability improvement process. The study shows, it is useful to obtain the availability importance measures value of each component in the system prior to deploying resources towards improving the specific components. With the assistance of importance measures one can identify the components/subsystems that merit additional research and development to improve their availabilities, so that the greatest gain is achieved in the system availability. Each component should be assigned a value and the component with a greater value will have a greater influence on the availability of the system. Generally, the importance of components should be used during the design or evaluation of systems to determine which components or subsystems have the greatest importance for the availability of the system.

Key Words: Availability, Availability importance measures, Failure rate, Repair rate, MTBF, MTTR

1 INTRODUCTION

It is clear that equipment availability continues to be a significant issue for industry. As the size and complexity of equipment continue to be increasing, the impacts of failure become even more critical. One method of mitigating the impact of failure is to improve reliability and availability of a system. Improvement of reliability and availability has been the subject of a large volume of research and a great number of articles have been published in this area [1, 2, and 3]. Availability and reliability are good evaluations of a system's performance. Their values depend on the system structure, as well as on the component availability and reliability. These values decrease as the component ages increase; i.e. their serving times are influenced by their interactions with each other, the applied maintenance policy and their environments [4]. The main requirements for the operation of complex systems are usually specified in terms of cost and availability and/or reliability, or equivalently in terms of mean operating time and/or mean down-time under a cost constraint. These requirements have then to be taken into consideration in the system design stage in order to determine the appropriate reliability and availability of each of the system's components [5].

In a simplistic sense there are some issues to be resolved during the development of an availability improvement or optimization process during design and operation, such as: i) where it is best to attempt improvements in availability, and ii) how to effect improvements in availability when the areas which merit attention have been identified. Finding appropriate answers to these questions, however, can be quite difficult and the solutions to the many problems which result in loss of availability are frequently not obvious. Reliability and availability engineers are often called upon to make decisions as to whether to improve a certain component or components in order to achieve better results. There are two ways to improve availability of repairable system: i) reduce the failure rate of component or components, and/or ii) improve the repair rate of component or components.

In this paper a methodology is defined to manage the availability improvement efforts of a repairable system, which consists of three steps. In order to manage the availability improvement efforts by applying those three steps, the concept of importance measure can be used, therefore some availability importance measures are defined which can be used as a guideline in process. The main objectives of this paper are: i) to define some availability importance measures in order to find the criticality of each components, and ii) to define a methodology for managing the availability improvement efforts.

2 SYSTEM AVAILABILITY DEFINITIONS

Availability is the probability that a system or component is performing its required function at a given point in time or over a stated period of time when operated and maintained in a prescribed manner [6]. Like reliability, availability is a probability. Consider a system (device) which can be in one of two states, namely 'up (on)' and 'down (off)'. By 'up' it is meant that the system is still functioning and by 'down' it is meant that that the system is not functioning. In the latter case the system is being repaired or replaced, depend on whether it is repairable or not. Let the state of the system be given by a binary variable:

$$X(t) = \begin{cases} 1, & \text{if the system is up at time t} \\ 0, & \text{otherwise} \end{cases}$$

An important characteristic of a repairable system is availability. Barlow & Proschan (1975a) [7] define four measures of availability performance: the availability function, limiting availability, the average availability function and limiting average availability. All of these measures are based on the function *X(t)*, which denotes the status of a repairable system at time *t*. The instant availability at time t (or point availability) is defined by:

$$A(t) = P(X(t) = 1) \tag{1}$$

This is the probability that the system is operational at time t. Because it is very difficult to obtain an explicit expression for A(t), other measures of availability have been proposed. One of these measures is the steady system availability (or steady-state availability or limiting availability) of a system, which is defined by:

$$A = Limit_{t\to\infty} A(t) = \frac{MTBF}{MTBF + MTTR} \tag{2}$$

This quantity is the probability that the system will be available after it has been run for a long time, and it is a very significant measure of the performance of a repairable system. If the failure rate and repair rate of system are constant, then the steady state availability can be calculated by equation 3.

$$A = \frac{\mu}{\lambda + \mu} \tag{3}$$

Consider a system which consists of *n* s-independent subsystems connected in series and which fails when at least one of its components fails. The steady-state availability for a series-system is the product of the component availabilities [6, 8].

$$A_s = \prod_{i=1}^{n} A_i = \prod_{i=1}^{n} \frac{MTBF_i}{MTBF_i + MTTR_i} \tag{4}$$

Consider a system which consists of *n* independent subsystems connected in parallel and which works when at least one of its components works. The steady-state availability of a parallel-system is given by equation 5 [6].

$$A_s = \coprod_{i=1}^{n} A_i = \coprod_{i=1}^{n} \frac{MTBF_i}{MTBF_i + MTTF_i} = \coprod_{i=1}^{n} \frac{\mu_i}{\mu_i + \lambda_i} = 1 - \prod_{i=1}^{n}\left(1 - \frac{MTBF_i}{MTBF_i + MTTFR_i}\right) \tag{5}$$

Consider a system which consists of *n* independent subsystems connected in series and each subsystem consists of m component, the steady-state availability for a series-parallel system is given by

$$A_s = \prod_{K=1}^{n}\left(1 - \prod_{l=1}^{m} 1 - A_{km}\right) \tag{6}$$

3 AVAILABILITY IMPORTANCE MEASURES

The concept of importance measures came from the perception that in any orderly arrangement of components in a system, some of the components are more important than others in providing certain system characteristics. In order to evaluate the importance of different aspects of a system, a set of importance measurements, including Structure Importance, Birnbaum Component Importance, Reliability Criticality Importance, Upgrading Function, Operational Criticality Importance, and Restore Critically Index [9, 10, 11] have been well defined and widely used in engineering practice. Birnbaum first introduced the concept of importance in 1969 and one of the most widely used reliability importance indices is Birnbaum's component importance [10]. The component reliability importance measure is defined as: probability that component i is critical to system failure [12]. The reliability importance of a component can be determined based on the failure characteristics of the component and its corresponding positioning in the system. The reliability importance measure of component i in a system of n components is given by

$$I_A^i = \frac{\partial R_s(t)}{\partial R_i(t)} \tag{7}$$

$R_S(t)$ is the system reliability, and Ri (t) is the component reliability.

It was found that the reliability importance of a component is a function of time and it is useful to obtain the reliability importance value of each component in the system prior to investing resources toward improving specific components. This is done to determine where to focus resources in order to achieve the most benefit from the improvement effort. If the reliability of the system needs to be improved, then efforts should first be concentrated on improving the reliability of the component that has the largest effect on reliability.

By using the same concept in the case of system availability performance, some importance measures can be defined and used as a guideline for developing an improvement strategy. One motive for applying this concept of importance measures in the availability improvement process is the fact that the concept is easy to understand and the applications of these measures are very simple. In this study the availability importance of a component is defined as a partial derivative of the system availability with respect to this component availability. Analyses of these measures for a numerical example are performed and the results are presented. Availability Importance Measures are function of operating time, the failure rate, the repair rate, and the system structure and it is useful to obtain the availability importance measure value of each component in the system prior to deploying resources toward improving the specific components. This is done to determine which component needs to be improved in order to achieve the maximum effect from the improvement effort. The availability importance of component i in a system of n components is given by

$$I_A^i = \frac{\partial A_s}{\partial A_i} \tag{8}$$

Availability importance measure shows that the effect of availability of component i on the availability of the whole system. The subsystem with the largest value has a greatest effect on the availability of the whole system. If the availability of the system needs to be improved, then efforts should first be concentrated on improving the subsystem that has the largest effect on the availability of system. The availability of a system is a function of Mean Time between Failures (MTBF)and Mean Time to Repair (MTTR). Therefore two other set of importance measures can be define in order to find to manage the availability improvement process for each component based on the MTBF or MTTR.

Availability importance measure based on the MTBF shows the effect of the MTBF of subsystem i on the availability of the whole system, and the subsystem with the largest value has the greatest effect on the availability of the whole system. It can be calculated by

$$I_{A,MTBF_i}^i = \frac{\partial A_s}{\partial A_i} \times \frac{\partial A}{\partial MTBF_i} \tag{9}$$

Availability importance measure based on the MTTR shows the effect of the repair rate of component i on the availability of the whole system, and the subsystem with the largest value has the greatest effect on the availability of the whole system. It can be calculated by

$$I_{A,MTTR_i}^i = -\frac{\partial A_s}{\partial A_i} \times \frac{\partial A}{\partial MTTR_i} \tag{10}$$

If the failure rate and repair rate are constant, then availability importance measure based on repair rate and availability importance measure based on repair rate can be used instead availability importance measure based on MTBR and availability importance measure based on MTTR.

$$I_{A,\lambda_i}^i = -\frac{\partial A_s}{\partial A_i} \times \frac{\partial A}{\partial \lambda_i} \tag{11}$$

$$I_{A,\mu_i}^i = \frac{\partial A_s}{\partial A_i} \times \frac{\partial A}{\partial \mu_i} \tag{12}$$

Comparing these two importance measures shows which of the two importance measures, the MTBF of The MTTR, has more influence on the availability of the whole system. In other words, this comparison will show whether the availability improvement should be based on reducing the MTTR or increasing the MTBF of critical components or subsystems.

3.1 Availability importance measures for a series system

The availability importance measures consider a series system in section 2 can be calculated by equation 13, 14 and 15.

$$I_A^i = \frac{\partial A_s}{\partial A_i} = \prod_{\substack{k=1 \\ k \neq i}}^{n} A_k \tag{13}$$

$$I_{A,MTBF_i}^i = \frac{\partial A_s}{\partial A_i}\frac{\partial A_i}{\partial MTBF_i} = \prod_{\substack{k=1 \\ k \neq i}}^{n} A_k \times \frac{MTTR_i}{MTBF_i(MTBF_i + MTTR_i)} \times A_i = A_s \times \frac{MTTR_i}{MTBF_i(MTTR_i + MTBF_i)} \tag{14}$$

$$I^i_{A,MTBF_i} = -\frac{\partial A_s}{\partial A_i}\frac{\partial A_i}{\partial MTTR_i} = \prod_{\substack{k=1 \\ k\neq i}}^{n} A_k \times \frac{1}{(MTBF_i + MTTR_i)} \times A_i = A_s \times \frac{1}{(MTTR_i + MTBF_i)} \tag{15}$$

3.2 Availability importance measures for a parallel system

The availability importance measures consider a parallel system in section 2 can be calculated by equation 16, 17 and 18. Equation 15 shows that the availability of a component doesn't affect on the availability importance measure of its component in a parallel system.

$$I^i_A = \frac{\partial A_s}{\partial A_i} = 1 - \prod_{\substack{k=1 \\ k\neq i}}^{n} A_k \tag{16}$$

$$I^i_{A,MTBF_i} = \frac{\partial A_s}{\partial A_i}\frac{\partial A_i}{\partial MTBF_i} = 1 - \prod_{\substack{k=1 \\ k\neq i}}^{n} A_k \times \frac{MTTR_i}{MTBF_i(MTBF_i + MTTR_i)} \times A_i \tag{17}$$

$$I^i_{A,MTTR_i} = -\frac{\partial A_s}{\partial A_i}\frac{\partial A_i}{\partial MTTR_i} = 1 - \prod_{\substack{k=1 \\ k\neq i}}^{n} A_k \times \frac{1}{(MTBF_i + MTTR_i)} \times A_i \tag{18}$$

3.3 Availability importance measures for a series-parallel system

The availability importance measures consider a parallel system in section 2 can be calculated by equation 17, 18, and 19.

$$I^{ij}_A = \frac{\partial A_s}{\partial A_{i,j}} = \prod_{\substack{k=1 \\ k\neq i}}^{n} \left(1 - \prod_{l=1}^{m}(1 - A_{kl})\right) \times \left(1 - \prod_{\substack{l=1 \\ l\neq j}}^{m} A_{il}\right) \tag{19}$$

$$I^{ij}_{A,MTBF_{ij}} = I^{ij}_A \times \frac{MTTR_{ij}}{MTBF_{ij}(MTBF_{ij} + MTTR_{ij})} \times A_{ij} \tag{20}$$

$$I^{ij}_{A,MTTR_{ij}} = I^{ij}_A \times \frac{1}{(MTBF_{ij} + MTTR_{ij})} \times A_{ij} \tag{21}$$

Equations 13, 16, and 19 are shown that the availability of a component doesn't affect on the availability importance measure of that component. In other words the availability importance measure of component depends on the availability of other component in the system and also the position of that component in system configuration.

4 AVAILABILITY IMPROVEMENT PROCESS

Availability is an important characteristic of a repairable system. When the availability of a system is low, efforts are needed to improve it. The question of how to meet an availability goal for a system arises when the estimated availability is inadequate. This then becomes a reliability and availability allocation problem at the component level. There are two ways to improve the availability of a repairable system, i) reduce the failure rate of the component in question or, in other words, increase the mean time between failures, and/or ii) improve the repair rate of the system, structure or component (SSC), or, in other words, reduce the mean down-time.

In the present research it is found that the availability improvement process could be implemented by following three steps:

Step 1: Identification of an ordered list of candidates for the availability improvement process.

Step 2: Identification of effective changes or remedial actions for each candidate, which will either reduce its failure frequency or reduce the time required to restore a component.

Step 3: Justification and prioritization of the actions for each candidate on the basis of cost-benefit comparisons.

In order to find the availability improvement strategy by applying these three steps, the concept of importance measures is used in this paper. One motive for applying this concept of importance measures in the availability improvement process is the fact that the concept is easy to understand and the applications of these measures are very simple. In step one an ordered list of candidates for availability improvement can be identified by using of the availability importance measure, but this measure does not provide more information about a component. Therefore, in step two the availability importance measure based on MTBF and the availability importance measures based of MTTR for each component must be used. Comparing the two importance measures shows which of the two factors, MTBF or the MTTR, has more influence on the availability of the whole system. In other words, this comparison will show whether the availability improvement should be based on reducing the failure rate or increasing the repair rate of critical components or subsystems.

To find the final strategy for the availability improvement process (step 3) the cost trade-off is essential. When the availability of the system is less, it needs to be improved by using the special budget C. The question is how to manage improvement efforts and which component or components, if improved, will give better results. This question can be answered through the following procedure. The cost needed to change the failure rate, denoted by $\Delta\lambda_i$, and the cost needed to change the repair rate, denoted by $\Delta\mu_i$, can be calculated by using equations 9 and 10.

$$\Delta C_{\lambda_i} = \frac{\partial C}{\partial \lambda_i} \times \Delta\lambda_i \tag{22}$$

$$\Delta C_{\mu_i} = \frac{\partial C}{\partial \mu_i} \times \Delta\mu_i \tag{23}$$

$\frac{\partial C}{\partial \lambda_i}$ and $\frac{\partial C}{\partial \mu_i}$ explain the variation of the availability improvement cost with respect to the failure rate and the repair rate of component i, respectively.

If budget C is spent on increasing the repair rate for the critical components, the repair rate will increase as $\Delta\mu_i$:

$$\Delta\mu_i = \frac{\Delta C_{\mu_i}}{\frac{\partial C}{\partial \mu_i}} = \frac{C}{\frac{\partial C}{\partial \mu_i}} \tag{24}$$

Therefore the availability will increase as $\Delta A_{s,\mu_i}$, which can be calculated by

$$\Delta A_{s,\mu_i} = I^i_{A,\mu_i} \times \Delta\mu_i = \frac{\partial A_s}{\partial \mu_i} \times \frac{C}{\frac{\partial C}{\partial \mu_i}} \tag{25}$$

If the budget is spent on reducing the failure rate of the critical component, the failure rate will be reduced as $\Delta\lambda_i$:

$$\Delta\lambda_i = \frac{\Delta C_{\lambda_i}}{\frac{\partial C}{\partial \lambda_i}} = \frac{C}{\frac{\partial C}{\partial \lambda_i}} \tag{26}$$

Therefore, the availability will be increased as $\Delta A_{s,\lambda_i}$, which can be calculated by:

$$\Delta A_{s,\lambda_i} = I^i_{A,\lambda_i} \times \Delta\lambda_i = \frac{\partial A_s}{\partial \lambda_i} \times \frac{C}{\frac{\partial C}{\partial \lambda_i}} \tag{27}$$

By comparing $\Delta A_{s,\lambda_i}$ and $\Delta A_{s,\mu_i}$ the strategy can be identified. If there are some restrictions, the budget can be spent on both increasing the repair rate and decreasing the failure rate. We then allocate a fraction *f* of the budget for decreasing the failure rate and the remaining fraction *1-f* for increasing the repair rate. And hence the availability improvement can be calculated by:

$$\Delta A_{s,\lambda_i,\mu_i} = \frac{\partial A_s}{\partial \lambda_i} \times \frac{fC}{\frac{\partial C}{\partial \lambda_i}} + \frac{\partial A_s}{\partial \mu_i} \times \frac{(1-f)C}{\frac{\partial C}{\partial \mu_i}} \tag{28}$$

5 CASE STUDY

To illustrate the model, we use a case study of a crushing plant in Bauxite mine of Iran. The crushing plant is divided into six subsystems and all subsystem work in series system which means the crushing plant is in working state if all subsystems work. The best fit distributions for all subsystem of the crushing plant are calculated by using Weibull ++6 based on historical data form the period of one year. Table 1 shows the best-fit distribution for Time between failures data and also for Time to Repair data for all subsystem of the crushing plants.

Table1. Best-fit distribution and thire parameters for TBF and TTR data sets

Sub-system	Time Between Failures		Time to Repair data	
	Best-Fit	Parameters	Best-Fit	Parameters
PCRCS (Primary Crusher)	Weibull 3 P	Beta=1.34, Eta=78.6,Gamma=3.873	Lognormal	Mean=0.4638, Std=0.922
SCRCS (Secondary	Weibull 3 P	Beta=1.115, Eta=78.96, Gamma=8.931	Lognormal	Mean=0.720, Std = 1.515
PSCCS (Primary Screen)	Lognormal	Mean= 3.373, Std = 1.1421	Weibull 2 P	Beta= 1.4998, Eta = 1.5843
SSCCS (Secondary Screen)	Lognormal	Mean= 3.868, Std = 1.101	Lognormal	Mean=0.10, Std = 1.021
COCS (Conveyer	Lognormal	Mean= 3.18, Std = 0.841	Lognormal	Mean=0.154, Std =1.1157
FECS (Feeder Subsystem)	Exponential 2P	Lambda=0.0057, Gamma= 24.80	Exponential 2P	Lambda=1.039, Gamma=0.159

Based on equation 13, 14, and 15, the availability importance measures for all subsystem of the crushing plant are calculated and tabulated in Table 2. The availability importance measure I_A^i shows that the SCRCS and COCS subsystems have more influence on the availability of the whole system, and, therefore, improvement in the availability of the SCRCS and COCS will cause the greatest increase in the system availability.

By comparing $I^i_{A,MTBF_i}$ and $I^i_{A,MTTR_i}$ one can determine whether the MTBF or MTTR has more influence on the availability of the crushing plant. In the case studied, if the availability of the crushing plant needs to be improved, one should first concentrate one's efforts on increasing the availability of the SCRCS and COCS. In addition, it is better to pay more attention to the MTTR of SCRCS and also MTTR of COCS subsystem, because the effect of MTTR for both of them, on the availability of the whole system is about 13 and 16 times respectively greater than the corresponding effect of the MTBF of both subsystem, which is indicated by a comparison of $I^i_{A,MTBF_i}$ and $I^i_{A,MTTR_i}$. However, the investment required to decrease the MTTR may be much grater than that required to increase the MTBF. Cost trade-off is essential for making final decision

Table 2. Availability importance measures for all Subsystem of the crushing plants

Sub-system	I_A^i	$I^i_{A,MTBF_i}$	$I^i_{A,MTTR_i}$
PCRCS	0.829	0.00033	0.01024
SCRCS	0.866	0.00068	0.00874
PSCCS	0.826	0.00040	0.01395
SSCCS	0.818	0.00016	0.00900
COCS	0.854	0.00139	0.02201
FECS	0.808	0.00002	0.00401

6 CONCLUSIONS

In this paper we propose a method for implementing the availability improvement process which consists of three steps. In the case of a system's availability performance, availability importance measures could be used as a guideline in developing a strategy for availability improvement. The availability importance of a component/subsystem is defined as a partial derivative of the system availability with respect to this component availability. It is useful to obtain the value of the availability importance measure for each component in the system prior to deploying resources toward improving the specific components.

7 REFERENCES

[1] Chales C & Kondo A. (2003) Availability allocation to repairable systems with genetic algorithms: a multi-objective formulation. *Reliability Engineering and System Safety*, Vol. 82, No. 3, December, p. 319-330

[2] Barabady J. (2005) Improvement of system availability using reliability and maintainability analysis, Licentiate thesis, Luleå University of Technology, Sweden.

[3] Richard C. C, Pohl E.A & Jin S. (2004) Managing availability improvement efforts with importance measures and optimization. *IMA Journal of Management Mathematics*, Vol. 15, No. 2, p. 161-174

[4] Samrout M, Yalaoui F, Chatelet E & Chebbo, N. (2005) New methods to minimize the preventive maintenance cost of series-parallel systems using ant colony optimization. *Reliability Engineering and System Safety*, Vol. 89, No. 3, p. 346-354

[5] Elegbede C &Adjallah K. (2003) Availability allocation to repairable systems with genetic algorithms: a multi-objective formulation, *Reliability Engineering and System Safety*, Vol. 82, No. 3, p. 319-330

[6] Ebeling C.E. (1997) An Introduction to Reliability and Maintainability Engineering. New York: Mc Graw-Hill

[7] Barlow R.E. & Proschan F. (1975a) *Statistical Theory of Reliability and Life Testing: Probability Models.* New York: Rinehart and Winston.

[8] Pham H. (2003) Steady-state series-system availability. *IEEE Transactions on Reliability*, Vol. 52, No. 2, p. 146-147

[9] Leemis, L.M. (1995) Reliability: Probabilistic Models and Statistical Methods. Englewood Cliffs, N.J.: Prentice Hall, Inc.

[10] Fricks M.F., Trivedi, K.S. (2003), "Importance analysis with Markov chains," *Proceedings of the 49th Annual Reliability and Maintainability Symposium*, p. 89-95

[11] Wang W, Loman, J& Vassiliou P. (2004) Reliability importance of components in a complex system. *Proceedings of the Annual Reliability and Maintainability Symposium*, International Symposium on Product Quality and Integrity, p. 6-11

[12] Beeson S &Andres J.D. (2003) Importance measures for non-coherent-system analysis. *IEEE Transactions on Reliability,* Vol. 52, No. 3, p. 301-310

[13] Barabady J. (2005) Reliability and maintainability analysis of crushing plants in Jajarm Bauxite Mine of Iran. In *Proceedings of the Annual Reliability and Maintainability Symposium,* USA, 2005, pp. 109-115

THE NATURE OF ENGINEERING MAINTENANCE WORK: A BLANK SPACE ON THE MAP

James Trevelyan, Leonie Gouws, and Adrian Stephan[a]

[a] *School of Mechanical Engineering, The University of Western Australia.*

Abstract: Engineering maintenance work is assumed and taken for granted. It is surprising to find that there are just a handful of carefully researched accounts of engineering work, and only one of engineering maintenance work. Yet there is strong evidence that engineering maintenance work is far from satisfactory with reported losses from equipment and plant breakdowns exceeding 40% of turnover. These losses often exceed direct maintenance costs by several times. While there are many prescriptions for organizing maintenance work, there is no research to substantiate these prescriptions. This paper reviews the available literature and describes research that is mapping the nature of engineering work in a form that could lead to significant improvements.

Key Words: engineering maintenance work, qualitative research, ethnography

1 MAINTENANCE IS CHEAP?

There is little doubt that many companies are continuing to have major problems with maintenance work and we outline just a few of the issues in this paper. Recent work has highlighted continuing high costs sustained by companies due to plant and equipment failures. Computer-based maintenance management systems (CMMS) have been proposed and implemented in the hope of controlling these failures and costs without much apparent effect. Given the direct costs of maintenance and also the magnitude of costs associated with maintenance failures, it is therefore surprising to find almost no research on the actual nature of maintenance work. The research literature contains little that helps us understand how it is done, the different roles played by different people, the skills and knowledge actually used. Instead the work is assumed to be known. After all, what could be so difficult to understand about some people carrying basic hand tools following directions on computer-printed slips of paper? Yet, ultimately, maintenance failures occur because people make mistakes or forget to perform important actions. These are people who experience meaningful contact with the machines or process plant and each other. We cannot hope to understand why maintenance fails to work well if we do not attempt to understand what these people are actually doing.

Recent work by Hägerby and Johannson [1] is just one of many accounts illustrating the magnitude of maintenance failures in modern enterprises. Figure 1 illustrates their results which show losses of up to 40% of turnover due to lack of plant availability. While this is a problem in industrialized countries, poor maintenance in developing countries leads to unacceptably high end-user costs for essential services such as water supplies. Poor people in developing countries simply cannot afford to wash their hands because the real cost of water is too high, typically up to 30 times higher than in Australia in real dollar terms. [2, 3]

2 RESEARCH LITERATURE

It is not easy to find the research that lies behind contemporary industrial maintenance texts [e.g. 4, 5]. Few include reference sources. After a thorough search we have found only a small number of research reports that provide any detailed insight on engineering maintenance work. Technical processes such as inventory controls, condition measurement, reliability forecasts and remediation methods have been reported in detail and it is tempting to assume that these methods describe what maintenance work actually involves. The few reports of systematic research, however, tell us that this assumption is likely to be false.

The first aim of this paper, therefore, is to review research literature that can inform us on the nature of engineering maintenance work. Even if we enlarge the scope of our literature search to include all engineering work, there is just a handful of systematically researched accounts that has appeared over the last 30 years.

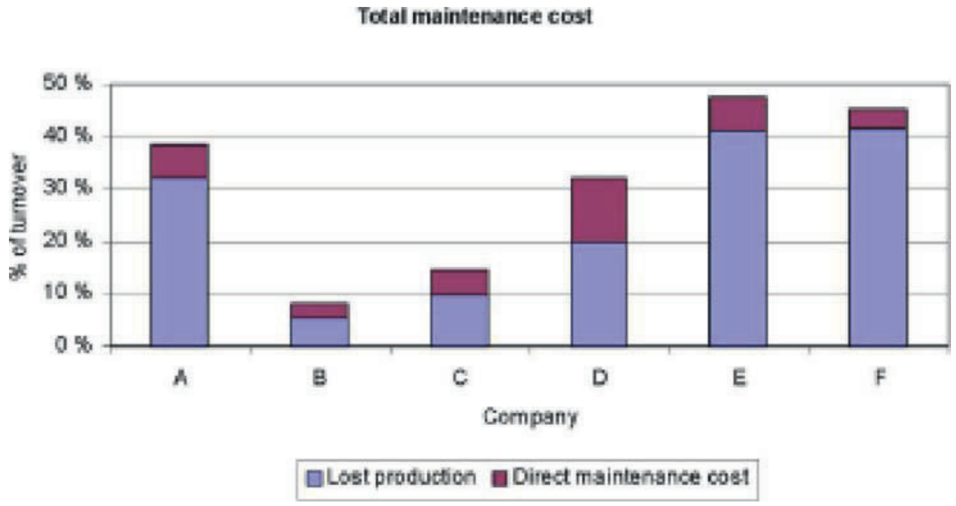


Figure 1: Results of survey of maintenance costs in six major Scandinavian firms. Note how the reported value of lost production is far higher than direct maintenance costs (reproduced from [1]).

What do we mean by ‘engineering work’? We would like to avoid distinctions between occupational groups, especially between ‘engineers’ as professionals and other workers such as technologists. Instead we will focus on all the work, including the planning and organization that leads up to the ultimate ‘production’ work whether executed by production workers, technicians, tradespeople or machines. It is essential to recognise that most engineering involves highly coordinated work by many people with different roles and success relies on the ultimate production work being done correctly. Engineers seldom do this themselves yet they are often held responsible for the production and service delivery work performed by other people, the ‘production’ or ‘service delivery’ workers. We can distinguish aspects of ‘engineering work’. Planning, analysis, design, organization, and administration precede the final ‘hands-on’ production or maintenance work that results in useful products, solutions and services, yet all are part of engineering work. In practice there is no clear dividing line between the different aspects. Engineers often engage in hands-on work and technicians are often involved in planning and coordination, as we see in Orr’s detailed account [6]. At the same time we also need to be interested in diverse roles of engineers. Engineering workers may refer to some of their roles as “not real engineering” work, sometimes “all that admin stuff”. They even organize social activities. Yet these roles are still part of the picture if we are to understand what engineering is all about.

In some discourse, engineering work is considered to be purely technical. For example, Hoag [7] presents a comprehensive work on continuing education and on-the-job learning for engineers without mentioning any non-technical skills training at all. On the other hand, Connelly [8] surveyed “personal and professional” skills, i.e. non-technical skills and knowledge, even though in the introduction they acknowledge that enhanced oil and gas recovery methods, decommissioning and abandonment will have an impact on the skills engineers need to acquire and develop.” Barley & Orr [9, Ch 2] provide evidence of a prevailing image that separates management from engineering. A person is either a manager or an engineer: the former has little or no technical content in his work and the latter has little or no social dimension to his work. However, our own work informs us that engineers and technologists perform roles that involve management of people.

There is a small group of richly detailed ethnographic style studies of working engineers, not all with the explicit intent of understanding engineering work. Several provide useful field data and observations which could be useful in testing different models of engineering work. However none seem to include a detailed classification of engineering work or a broad enough coverage to attempt to build one. The few studies informed with detailed technical understanding are mainly on design, only one of many roles undertaken by working engineers. However design carries considerable glamour and status, particularly in high technology and aerospace, and this area of engineering work has been studied extensively by many researchers. However, for most of them their interest lies in the *design process* rather than all the other less glamorous work that designers also have to accomplish (For example, see Eckert [10] for a description of recent work).

A recent survey of about 100 research publications [11] reveals two major groups of reports. Investigations on engineering education have stimulated about half of all the reports on engineering work. The other half has emerged from researchers who have been interested in engineering management. However, the vast majority of these results were obtained using methods based on questionable assumptions. Only a small number can be relied on, mostly using anthropological approaches based on qualitative research. Most of these were written by social scientists or students with little engineering experience.

Schön [12] is one of the earliest accounts of this type. His interests lie in the reflective aspects of professional work and he describes engineers inventing a new process. He uses the term "technical rationality" to describe science-based professionals as technical problem solvers. According to this model "problem-solving is a manipulation of available techniques to achieve chosen ends in the face of manageable constraints."

Solomon and Holt [13] use an ethnographic approach to study mechanical engineering work with the stated aim of revisions to engineering education. The paper appears to foreshadow a more extensive analysis though none seems to have subsequently appeared.

The work of Bucciarelli [14, 15] is significant because he is one of the few engineers who has made extensive contributions to the literature on engineering work. He introduces the useful notion of “object worlds” – domains in which engineers conceive their designs and explain how they work, analogous to the “planes of objects” described by Foucault [16].

In their collection of studies, Barley & Orr [9] attempted to define technical work. They suggested that such work "requires understanding and utilisation of abstract knowledge", presumably mathematics or science acquired through formal education. They acknowledged that many technical workers, particularly technicians and trades workers, acquire their knowledge and skills through experience and working with more experienced workers (e.g. apprenticeships), rather than through formal education. They also suggested that “technical acumen is the result of contextual and tacit knowledge”. "Technicians are usually charged with ensuring that machines, organisms and other physical systems remain in good working order. Caretaking often requires technicians to employ theories, diagnoses, documentation, and other representations drawn from the symbolic realm they support.” [17, p89]

Barley and his co-workers have made several other significant contributions. Barley and Bechkey [17] point out that technicians mediate between professionals and the phenomena over which the professionals are reputed to have mastery. To do this technicians translate "aspects of physical phenomena into science, symbols, or information that physicians, scientists, engineers and other professionals subsequently use in their own work". "Technicians in nuclear power plants... create and monitor flows of data on production systems that engineers and managers used to make decisions." They also report

"informants claimed that accomplished technicians were obsessively organized.... lab workers spoke of maintaining records, scheduling activities, and keeping supplies in their proper place.... lab staff viewed 'being organized' as the primary weapon in their war against uncertainty. To say that a technician was highly organized was to say he or she was adept at a work style that lab workers viewed as independent of procedure or discipline."

Orr [6] also reports this in his highly detailed ethnography of photocopier technicians working in a West Coast USA commercial environment. His contribution is soundly based on extensive first-hand technical knowledge through his previous career as a maintenance technician. He concludes that chaotic work is associated with finding and solving problems by luck rather than by competence and is frowned upon in the community of maintenance technicians. This is the only report we are aware of that presents a comprehensive account of engineering maintenance work. Orr describes relationships between service technicians, customers and individual machines. While the work is very accessible in its writing style, there is no attempt to derive a systematic description of the work. Instead, the focus is on the technicians as a group and a description of their social networks with an anthropological audience in mind.

Several other social scientists have studied engineering work though they have been mainly interested in the place this occupies in society and firms rather than the detail of the work itself [See for example 18, 19-23]. Social scientists have acknowledged difficulties in understanding technical work. "Engineers continue to play havoc with conventional models of work even though they are the prototypical technical workers of high industrialisation." In this one sentence the authors admit that they really do not understand the work of engineers. [20, p34]. One of the obvious barriers for social scientists is technical knowledge, presenting an intellectual dimension that requires years of practice to fully comprehend. “... much of engineering is desk work based on thought processes whose character is manifest only partially in the various pieces of paper that are intermediate product. ... Some aspects of engineering work, especially its technical content, resist observation…” [21, p28-29, also p35].

One of the more recent contributions has been a collection of work by Vinck [24] and his colleagues. In the first chapter we find a detailed account of the work of a young engineering intern as a way of gaining a better understanding of engineering design work. Once again the focus is on design work, but there are some highly detailed accounts of various engineering roles which can provide useful evidence for testing models.

Ergonomics and human factors engineering researchers have studied specific work practices, of course, but the focus has predominantly been on specific machine and process operators, for example, pilots and air traffic controllers. [25] provides a useful example: a detailed analysis of errors in aircraft maintenance work, yet a detailed description of the work seems to be assumed: it is not described.

Vinck has also pointed out how weak our real understanding of engineering and technological work really is.

“We live in an era of paradoxes. On the one hand, we are faced with an ever-growing array of technological developments that affect communication, work, travel, domestic and leisure activities, and political and ethical debates. … On the other hand, in many ways we do not understand these technologies. The surprising thing is that this is true for technological professionals as well as for lay people. Our situation is characterised both by our ignorance of new technologies and by our faith in them." [24]

It seems that our knowledge of the nature of most engineering work, including maintenance work, remains largely at the state that our knowledge of management work was in the mid 1960s when Mintzberg [26] started his research on the nature of managerial work. The questions he posed then in relation to managers could just as well be asked about engineers and technicians today.

3 MAP MAKING

In this section of the paper we will describe our ongoing research that aims to fill this blank space on the map. It is part of a larger project on the nature of engineering work using contemporary social sciences investigation methods. This project has already yielded some valuable insights from preliminary results coming from analysis of 55 interviews with practising engineers, drafting staff, technicians, tradespeople, planners and schedulers with similar fieldwork in several countries. Most of the interviews, so far, have been with engineers.

We have used an ethnographic research technique. We analyze transcripts taken from 90 minute semi-structured interviews that explore the daily work of an individual involved in engineering work. The analysis extracts quotations that provide evidence for different kinds of engineering work roles. We use three 'triangulation' methods to check the validity of this analysis. First, the interviews and analysis are performed by experienced engineers. This has proved to be essential because lack of significant engineering experience makes it difficult to understand what can lie behind a cryptic comment by an engineer or technician. Second, we shadow a limited number of interview subjects, recording their activities over a limited period ranging from 1 day to a week. By doing this we can check for the relative significance of the interview comments and check for roles missed in the interview. Through this process we have gradually built up a relatively complex framework of engineering roles and technical knowledge categories, defined in terms that are not specific to any one engineering discipline (see Appendices for examples). The third triangulation method is to check the limited available research literature to see whether we have missed roles and knowledge categories, and also to see whether our results can explain observations made by other researchers.

One result is a systematic framework for interpreting the nature of engineering work [27]. The framework consists of approximately 80 different roles performed by engineers and other people involved in engineering work (for a small selection see Appendix 1). Certain roles have dominated interview responses, particularly roles that involve the coordination of work by other people over whom engineers have little or no authority. This can only happen with cooperation and our interviews show that an important feature of engineering work is gaining the *willing* cooperation of other people. While similarly non-technical generic roles are more prominent in interview responses than purely technical roles, it is clear that technical knowledge still plays a vital part. However, the kinds of technical knowledge and the ways in which it is used appear to be only distantly related to the technical content of university engineering training. This helps to explain why social scientists find it difficult to make progress in understanding the nature of engineering work: understanding the technical dimension is crucial.

An examination of the technical knowledge used in engineering work reveals another factor that helps to explain difficulties in maintenance work. We identified about 35 categories of technical knowledge ranging from knowledge of components and suppliers to process applications. Some of these categories appear in appendix 2. Our research shows that much of the knowledge used in engineering work is unwritten knowledge and people are often not consciously aware of the knowledge that they use on a daily basis. This is often referred to as tacit or implicit knowledge [28].

Another result concerns the ways in which people interact and communicate in the course of engineering work. These results might help to explain many of the difficulties encountered with maintenance management information systems. Oral or spoken interactions seem to be more important in many ways than written forms of communication, particularly in the effective coordination of technical work.

From our interviews there are several dimensions to technical work and different people appear to have different strengths in each.

- Familiarity with contextual and tacit technical knowledge usually acquired in a framework provided by formal education typically covering mathematics and science subjects.
- reative thinking, ability to conceive new approaches to technical problems that resist logical and analytical solutions.
- Manual skills acquired through a combination of childhood play, curiosity with respect to mechanisms and machines, possibly classes and formal education and finally on the job or extracurricular practice.
- Social skills that involve interactions with other people to obtain information and cooperation in technical work.
- Ability to manipulate abstract concepts, with a particular affinity for mathematics and philosophy that would be almost certainly strongly reinforced by formal education.
- Ability to pay meticulous attention to details that enables the person to work systematically through complex processes with few if any errors.

All five of these dimensions seem to be present in engineering workers to varying extents. Many prefer analytical and technical work, almost seeking refuge from personal interactions by burying themselves in this kind of occupation. Others have strong social skills and in some cases show signs of atrophy in their technical and analytical skills. Many kinds of engineering work require meticulous attention to detail without necessarily a strong abstract theoretical and analytical component. Many engineers have negligible manual skills: mechanical engineers are often notorious for the poor mechanical state of their cars. However a minority have highly developed manual skills and can outperform highly skilled technicians. Some engineers adapt to management very easily because they have highly developed social skills. Others with less advanced social skills manage but with much greater difficulty and less performance. There is an intriguing contrast in the apparent characteristics of some of the people involved in engineering asset management and maintenance. Asset management analysis, and particularly the use of computer-based asset management systems requires someone who feels comfortable with extensive and detailed written documentation. Meticulous attention to detail is essential for success. Yet maintenance management requires the successful coordination of in-house technical staff, operations staff, logistics, and often a host of outside contractors. 'Firefighting' is often used by these people to describe their role and they have little time for patient analytical work. The maintenance work itself is performed by technicians who need to have well developed manual skills and the research reports we have suggest they exchange much of their technical information verbally. Finally operations personnel also interact with the machinery and they seem to have a strongly developed verbal culture. Written information, even if available, is typically incomplete and seldom referred to directly. This explains communication difficulties that we have observed between these different people, especially if a CMMS forms a significant component of the communication process.

In a companion paper [29] we have noted the relatively low status of engineering maintenance work. One significant result from our interviews with engineers across a wide spectrum of disciplines is that maintenance was only mentioned by a tiny minority. From this we have confirmed that maintenance work is not a significant part of work for many engineers.

Social science methods, particularly ethnographic and qualitative research methods, provide some powerful research tools to investigate maintenance work practices. However they are expensive. We have learned that experience in engineering work is a pre-requisite to understand the technical factors that underlie the work. It takes time for people with an engineering background to learn the required investigation techniques, develop interviewing and field observation skills, read the required literature, and understand precautions that have to be taken to avoid personal bias affecting the results. The results come only through writing and distilling huge amounts of written text in the form of interview transcripts and field observation notes into readable 'quotes' and 'vignettes' that illustrate the essential aspects of working roles. Once a robust theoretical framework can be established with these methods, we can develop quantitative survey questionnaires that will help collect data from a wider range of workplaces.

All this takes time. It will take several years to collect sufficient data to provide a comprehensive analysis of engineering maintenance work. However, there are many potential benefits, some of which could be obtained without waiting so long. With a clearer understanding of how knowledge is transferred in engineering work we can already be confident that direct human contact is necessary to retain the value of unwritten knowledge that cannot be represented in a CMMS. Similarly, supervision and control of maintenance work is likely to be more effective with regular first-hand exposure to the technical work. History data is likely to be more accurate if it is checked when first entered by engineers who are likely to use the data rather than relying on maintenance technicians. By clearly understanding maintenance work we can define more effective training and development programmes. While this approach to maintenance requires more costly oversight, it may provide large savings in indirect maintenance costs.

4 CONCLUSIONS

Engineers often complain that their work is poorly understood by outsiders. In contrast to architects, lawyers, accountants, doctors and preachers, engineers have little face-to-face contact with the general public. Even if senior non-technical staff in major companies were curious enough to learn more about what their engineers do, there are almost no easily accessible accounts that they could read. The work of technicians is mainly invisible and taken for granted, even by engineers: it only becomes obvious in its absence or failure. A comprehensive description of engineering work, particularly engineering maintenance work, could be of enormous help.

The analysis and research necessary to compile such a description could also lead eventually to significant improvements in work practices. While management work has been researched systematically for more than a century, it is only in the last 40 years that our understanding has been sufficient to provide new insights, working methods and training programmes. Research on the nature of engineering work has hardly begun.

5 REFERENCES

[1] M. Hägerby and M. Johansson, Maintenance Performance Assessment: Strategies and Indicators, in Department of Production Economics. 2002, Linköping Institute of Technology: Linköping. p. 177.

[2] J. Trevelyan, Drinking Water Costs in Pakistan. 2005: School of Mechanical Engineering, The University of Western Australia. http://www.mech.uwa.edu.au/jpt/pes.html

[3] J. Trevelyan, Pakistan: Expensive Labour and High Costs. 2005: School of Mechanical Engineering, The University of Western Australia. http://www.mech.uwa.edu.au/jpt/pes.html

[4] R. Smith and B. Hawkins, Lean Maintenance. 2004, Burlington, USA: Elsevier Butterworth Heinemann. 290.

[5] S. J. Thomas, Improving maintenance and reliability through cultural change. 2005, New York: Industrial Press. 356.

[6] J. Orr, Talking About Machines: An Ethnography of a Modern Job. Collection on Technology and Work, ed. S. Barley. 1996, Ithaca, New York: Cornell University Press. 172.

[7] K. Hoag, Skills development for engineers: an innovative model for advanced learning in the workplace. IEE Management of Technology Series 20, ed. J. Lorriman and G. A. Montgomerie. 2001, London: IEE. 122.

[8] J. D. Connelly and J. C. R. Middleton, Personal and professional skills for engineers: one industry's perspective. *Engineering Science and Education Journal*, 1996(June): p. 139 -142.

[9] S. Barley and J. Orr, eds. *Between Craft and Science: Technical Work in US Settings*. 1997, Cornell University Press: Ithaca, New York.

[10] C. Eckert, A. Blackwell, L. Bucciarelli, J. Clarkson, C. Earl, T. Knight, S. Macmillan, M. Stacey and D. Whitney. What designers think we need to know about their processes: Early results from a comparative study. in *International Design Conference - DESIGN 2004*. 2004. Dubrovnik.

[11] S. Tilli and J. Trevelyan, Published Research on the Nature of Engineering Work. 2005: School of Mechanical Engineering, The University of Western Australia. http://www.mech.uwa.edu.au/jpt/pes.html

[12] D. A. Schön, The Reflective Practitioner: How Professionals Think in Action. 1983: Basic Books Inc., Harper Collins. 370.

[13] F. L. Solomon and J. E. Holt, On the Nature of Mechanical Engineering Work: An Analysis of Practice. *International Journal of Engineering Education*, 1993. **9**(6): p. 442-452.

[14] L. L. Bucciarelli, An ethnographic perspective on engineering design. *Design Studies*, 1988. **9**: p. 159-168.

[15] L. L. Bucciarelli, Designing Engineers. Inside Technology, ed. W. Bijker, B. Carlson and T. J. Pinch. 1994, Cambridge, Massachusetts: MIT Press. 220.

[16] M. Foucault, The Order of Discourse: Innaugual lecture at College de France December 1970, in *Language and Politics*, M. Shapiro, Editor. 1984, Blackwell: Oxford. p. pp108-138.

[17] S. Barley and B. A. Bechkey, In the Backrooms of Science: the Work of Technicians in Science Labs. *Work and Occupations*, 1994. **21**(1): p. 85-126.

[18] S. Crawford, Technical Workers in an Advanced Society: The Work, Careers and Politics of French Engineers. 1989, Cambridge: Cambridge University Press.

[19] P. Whalley, The Social Production of Technical Work: The Case of British Engineers. 1986: MacMillan.

[20] P. Whalley and S. Barley, Technical work in the division of labour: stalking the wily anomaly, in *Between Craft and Science: Technical Work in US Settings*, S. Barley and J. Orr, Editors. 1997, Cornell University Press: Ithaca, New York. p. 23-36.

[21] R. Zussman, Mechanics of the Middle Class: Work and Politics Among American Engineers. 1985, Berkeley: University of California Press.

[22] P. Meiksins and C. Smith, Engineering Labour: Technical Workers in Comparative Perspective. 1996, London: Verso.

[23] J. K. Fletcher, Disappearing Acts: Gender, Power, and Relational Practice at Work. 1999, Cambridge, Massachusetts: MIT Press.

[24] D. Vinck, ed. *Everyday Engineering: An Ethnography of Design and Innovation*. Inside Technology, ed. W. Bijker, B. Carlson and T. J. Pinch. 2003, MIT Press: Boston. 247.

[25] J. Reason and A. Hobbs, Managing Maintenance Error. 2003: Ashgate. 183.

[26] H. Mintzberg, The Nature of Managerial Work. 1973, Harper and Rowe: New York.

[27] J. P. Trevelyan, Definition of Engineering Roles. 2005, Perth: School of Mechanical Engineering, The University of Western Australia. http://www.mech.uwa.edu.au/jpt/pes.html

[28] M. Polyani, Personal knowledge: towards a post-critical philosophy. 1962, New York: Harper Torchbooks.

[29] L. Gouws and J. P. Trevelyan. Research on influences on maintenance management effectiveness. in 2006 World Conference on Engineering Asset Management. 2006. Surfer's Paradise, Queensland: Maintenance Engineering Society of Australia.

6 APPENDIX 1

This selection describes 17 of the 80 roles we have defined in our model of engineering work. Not all people perform all roles: engineering teams bring people together to provide expertise needed for all the roles required in a given project.

Personal research	Research technology developments, organisational and planning methods, business approaches, components and alternatives: when one type of material or component is not available which ones can be used instead, all to build personal knowledge and expertise.
Apply for job	Apply for positions outside organization, register with employment agencies, maintain current resumé and portfolio of achievements.
Documentation	Document existing systems, components, processes, conduct tests where necessary to determine performance, material properties etc.
Respond to query	Respond to technical, business or financial questions from peers, subordinates, sub-contractors and clients on construction, installation, maintenance and operating procedures. Suggest repair and maintenance procedures when unexpected operating conditions result in failure to achieve required results.
Acceptance test	Perform (or supervise) acceptance or commissioning tests on purchased (or in-house produced and/or installed) equipment, software, systems and materials, to ensure compliance with specifications.
Diagnose problem	Diagnose performance issues (may be partial or total failure, or sub-optimal or sub-specification performance), check with databases, consult experts, gather appropriate evidence, perform measurements, locate the root cause, and possibly devise alternative technical solutions.
Predict failure modes	Predict component and system failure modes and probabilities, determine consequences and costs of failure
Work package instructions	Design and write instructions, construction, installation, maintenance and operating procedures and instructions, preparing work packages.
Assign tech staff	Allocate responsibility for technical work: balance technical expertise and experience against cost and availability, decide whether to employ additional staff or contractors etc. Select appropriate working methods and tools.
Plan maintenance	Plan maintenance, condition monitoring, maintenance and asset management plan for components and systems with high failure costs and/or consequences
Coordinate insiders	Supervise or coordinate work of peers, subordinates and superiors. Perform technical checks on work, watch for roadblocks, may provide advice and feedback, may review technical competence, may assess training needs, provide informal training when appropriate.
Recruit staff	Assess qualifications, knowledge, skills and experience of engineers and technical staff from resumé, curriculum vitae, referees and employment reports for a given job description. Interview engineering and technical/administrative job applicants for given job description, assess suitability.
Hands-on tech work	Hands-On technical work; production or maintenance
Follow org'n procedures	Learn and practise standard organizational procedures (eg purchasing, tender preparation, design review, configuration management, safety risk assessment, quality measurement, quality control and quality management, etc.)
Manage production	Manage production capacity and resources for product delivery (product is the output of resources being managed: can be people, plant or machines or combination).
Invoicing & debt collection	Manage invoicing and debt collection, payments to contractors, maintain background knowledge on contractor financial status.
Networking	Networking; develop and maintain network of contacts to help with performance of job

7 APPENDIX 2

Our framework of technical knowledge is built on the notion that engineering creates products and services by combining and transforming components and materials with predictable value, performance, cost and timescale so that the people involved can be adequately remunerated and investors rewarded with appropriate return on capital. Most of the materials and components are usually the results of other engineering activities.

The following list is a selection of technical knowledge categories applicable in typical engineering disciplines. Each discipline (or even organization) will involve different items under these categories. Some categories may involve thousands of separate items, for example, knowledge of components and materials. Much of the working knowledge is unwritten knowledge, sometimes known as 'know-how'.

Applications of the product: knowledge of how product is applied and used by client's personnel. The client is more likely to have detailed knowledge than the engineers developing and making the product.

Customer needs (technical, social, business): understanding of individual client needs, including technical requirements for performance, life, maintainability, compatibility, industry standards etc., understanding client social needs, both within local community and as individual people, understanding client business needs such as price, financing method, delivery timing, client organisation logistics, business rules, installation, maintenance, service and support, repairs etc.

Definition of Product: knowledge of product details and components, function and location of each component, how each component contributes to the function of the whole assembly.

Diagnosis methods for product: technical methods for collecting data and analysing data to determine the cause of performance loss or failure. This also relies on having a model of how the product functions (T01) and relevant physical principles (T29).

Properties of completed product: models of product behaviour (technical and commercial) in given applications. Software, mathematical models of technical or commercial performance. Methods for predicting product performance, both commercial and technical.

Prediction, forecasting: knowledge to predict future events based on present conditions, anticipation of future events, knowledge of models and methods to calculate future conditions.

Industry standards and codes applicable to product and components, government regulations, including national standards (e.g. BS, DIN, AS), industry standards (IEC, API), and international standards (ISO), codes for design and operations (ASME, ASCE, ASTM etc.), government regulations for technical work. Companies can also have their own internal standards.

Measurement, test, inspection methods for product and components

Normal standards for technical work, production faults and defects in product and components, symptoms of "trouble"

Manufacturing methods, assembly of product, construction methods, time/cost and resources needed

Control of production or working environment, elimination of errors, quality control, safety.

Failure symptoms, signs of "trouble" with the product: problem indicators, clues, queues and symptoms can include unusual noise, smell, heat, and visible fragments (chips, smoke, rubbings, flakes, fragments, dust, particles -- particles may be attracted or retained by secondary effects such as wet or oily surfaces). Evidence of heating can include discolouration, burn marks, heat haze, shimmering as well as direct tactile contact or temperature indications. Visible damage can include cracks, dents, scratches, marks, rubbing, bends, crazing, discoloration, wrinkling, distortion in light reflections, and burn marks. With electronic equipment signs of trouble can include intermittent failures, heating, even audible noises. Problems can also be evident from operator comments, even the contents of trash bins, for example discarded copies from a photocopier [6].

Knowledge of components, materials

Information retrieval: locating required technical information in large amounts of mostly irrelevant written documentation.

Defining complete list of parts and materials required for given assembly and tools and other equipment required for construction, production and assembly.

Human behaviour, in design, production work, maintenance and operation of the product.

THE FUTURE OF ASSET MANAGEMENT

Author: S. Bradley Peterson

Presenter: Ralph Hedding

Strategic Asset Management Inc.

Abstract: Deming was famous for asking people, "By What Method?" an objective would be accomplished. Any useful concept has embedded in it the answer to that question.

Asset Management, to be a valuable concept, must demonstrate a method that leads to success. The method should be inclusive of all the necessary elements, yet be simple to understand.

We present a broader view of Asset Management here, built upon three core operating functions: Lead, Execute and Enable. The Execute function is always performed in the plant environment. Manufacturing success is relative to the competition however. Whoever can make the best products at the lowest cost will win. Lead and Enable are the great determinants of that relative success.

In addition, Physical Asset Management has an identity crisis. Most practitioners use the term synonymously with the maintenance function. Asset Management's opportunity, however, goes beyond maintenance to be a guiding principal for all of manufacturing. This paper explores the four integrated facets of the Execute function in the plant:

Capacity acquisition and disposition

Production

Asset Healthcare

Logistics

Our model, then, represents the Future of Asset Management, suggesting "The Method", necessary elements for successful manufacturing.

Key Words: Asset Management, Maintenance, Function, Manufacturing, Execute ,Healthcare, Logistics.

1 INTRODUCTION

Any useful model to guide action will have several characteristics.

- Simplicity. All of the greatest ideas are simple in concept. If not kept simple, they are not yet fully understood or remembered, and fail as guiding principles.
- Intuitive. Readers should be able to understand the underlying principles without guidance.
- Utility. The model should work consistently in application.
- Completeness. All necessary elements of success should be contained.

Our experience with the ***Strategic Asset Management*** (SAM) model (Figure 1) indicates these criteria are met. However, you as the reader will make your own judgments.

Our description of *Strategic Asset Management (SAM)* begins with the key elements of success, namely ***Lead, Execute and Enable.***

2 RECOMMENDATIONS

2.1 LEAD.

There are hundreds of books written about leadership, and each one has some valuable point to make. Frequently, though, the description is of individual leadership, and often charismatic leadership. Leadership in the plant setting, in our opinion, is creating consistency of purpose and action. Manufacturing is a large set of complex and interrelated systems of marketing, technology, finance, human resources, execution functions and equipment. Physical Asset Management then must take all of these into account.

Putting things as simply as we can into the SAM model, the LEAD element consists of the Managing System (Figure 2), Strategic Planning and Information Management.

Managing System. Disciplined, aligned action is the underpinning of any human endeavor. Examine any consistently great achievement and you will find alignment and discipline. That is the purpose of the managing system. Among the elements found here are:

- Top down and cascaded goals. Goals of profitability at the company level become volume and product mix goals for the plant. At the unit level these become volume goals, equating to equipment availability and product quality goals. For the operator these become daily production and equipment surveillance goals. For the craftsman they become equipment condition goals.
- Plan-Do-Review. Everything we do is a process. Take emergency maintenance for instance. The process is: 1. get a request, 2. go do it, repeated all day long. Even a planned and scheduled job doesn't improve the system without a review process to examine the effectiveness of the plan, the execution of results, and a critical understanding of what is happening with the equipment.
- Measurement systems. Assuring that in addition to outcome (lagging) indicators, each job in the plant has process (leading) indicators will enable each worker to make a more positive contribution.
- Reward systems. Each plant rewards behavior in subtle ways. It may promote the overbearing craftsman to a supervisor position because "he gets people going". We may reward equipment breakdown with money and admiration (overtime and attaboys). We may reward production achievement at any cost to people and asset condition. We reap what we sow, so we must be careful that our reward systems actually encourage proactive behavior.
- Roles, responsibilities and accountabilities are clear. If job expectations are not clear and results measurable we have muddled accountability. Fingers point in all directions, and the blame game goes on all day. Being proactive in such a system takes more courage than most people will risk. Clarifying jobs and accountabilities is a leadership function.
- Feedback. This is part of the Plan-Do-Review process, but gets special emphasis. We shape behavior by giving honest feedback without punishment. Under the right circumstances people *want to improve*. Leadership fails if it doesn't capture that spirit.

Strategic Planning. In every plant environment we encounter we hear the same (legitimate) complaints: "Improving maintenance is important, but we just don't have time. We have 4 major plant initiatives and 5 corporate initiatives, and don't know how any of them are going to get done!" Or, "Everything we do is a 'flavor of the month'. We seem to start lots of stuff, but never finish".

How do you set a plan that is timeless? One that has approvals all the way to the top of the company? How do you assure you have a single initiative instead of 20? **Through the act of *strategic planning!***

The product of functional strategic planning is *alignment around a multi-year improvement plan.* To get alignment requires more than a few words in a book. It requires that every level of the organization believes the content of the plan is the most important set of things the company can do with its resources. That means a real and compelling business case for the senior executives. For plant executives it means working on those things which are most practical, which make a difference in daily control of the work and reduction of variance. For the staff functions it means an understanding of the support they must render to enable the plan to be successful.

The elements of the Strategic Plan are these:

1. Benchmarking the function. Where are we today? What are the measures saying?
2. Developing a vision for the future of plant operations. Difficult to do, sometimes requires “Industrial tourism” to see the bigger picture, and using outside help to understand what’s possible. This part has to be done right, or the plan will fall apart. Our clients have found the Execution Triangles (Production, Logistics and Asset Healthcare) to be valuable to setting the vision.
3. Identify Gaps. Where do we fall short of the vision?
4. Identify Strategies to close gaps. It would be easy to shortcut this task. But one may find that, for instance, a distributed control system may be a strategy that helps with product quality, product mix direction, faster changeovers, and equipment condition monitoring. So one strategy cover several gaps.
5. Describe Projects to implement strategies. This can get creative, and will be an integrating force. For instance, a planning and scheduling project may be combined with a safety improvement initiative. Preventive maintenance improvement may combine with and ISO calibration standard.
6. Develop the implementation plan. It will require resources, so don’t shortcut or lowball what the implementation will require. Remember that training won’t create new behaviors. People need to be coached for new behaviors, and measures put in place to determine success.
7. Develop the Business Case. By integrating the initiatives into a single Strategic Plan we can avoid the silliness of double counting for results. Was contractor reduction due to the purchasing initiative or planning and scheduling? Who cares? As long as the goals for contractor reduction were met, and we stayed within the resource guidelines we requested and received approvals for.
8. Create the implementation governance. Plant leadership integrates the Strategic Plan into the Annual Planning Cycle, and the entire Managing System is engaged to see that the results of the Strategic Plan have accountabilities built into the entire organization.

Information Management. The good and bad news: as of the end of the last century most plants are now working with an ERP system. Initial results are typically very negative: lots of rejection of the new system as hard to use, can’t get reports out. But slowly organizations learn to live with and even like the new systems.

The deficiency, as always in IT, is there is confusion as to the difference between the *system* and the *tool.* The system is your set of internal processes and procedures. The tool may be SAP PM module. When your actual work process and methods aren’t reflected in your tool, the disconnect creates great dissatisfaction and waste. But when integrated, there is great synergy to get information to manage the business.

After the Strategic Plan, we undertake a detailed design for one or more of the Execution functions. In that design process we assure complete alignment between the tool and work processes, leading to a virtuous cycle of increasing understanding and utility of the system.

2.2 EXECUTE. Four functional areas exist in any manufacturing environment. These arenas are the typical focus of leadership. If done well, they lead to *functional excellence.*

- Capacity Development is usually considered to be the Design Engineering and Project Management functions in an organization. This function consumes $100’s of millions in what are often risky bets made on optimum market assumptions. A thoughtful and complete method to assure all excellence in the assumptions, design, construction and preparation for production can be a valuable tool. This is a future area for SAMI to define.

- Production Management. Everyone in the plant believes with good reason that production is the reason we are all here. And indeed this is the vehicle for value creation.
- Logistics include materials management, purchasing and movements of people and materials. This function can make or break the Production and Asset Healthcare Management functions.
- Asset Healthcare Management. Is this just another term for maintenance and reliability? Perhaps at some levels it is. But it is concerned with optimizing and integrating with all parts of the business based on risk and value, and so goes beyond the traditional boundaries of M&R.

We have developed improvement models for Production (Figure 3 & 4), Logistics (Figure 5 & 6) and Asset Healthcare (Figure 7 & 8). Indeed the Asset Healthcare model is the well-known *SAMI Triangle*, relabeled and integrated in the context of *Strategic Asset Management.*

2.3 ENABLE.

Many programs for change are viewed as a simple matter of documenting procedures and providing training. If these things are done, change should happen, right?

Wrong. Human nature doesn't work that way. Prescriptive formulations may work for machinery, but (fortunately) the human machine is more complex!

First, why do I say this is fortunate? Most of our corporate clients have a large number of related but disintegrated goals. Safety, reliability improvements, 6 Sigma, Lean Manufacturing, Supply Chain rationalization, etc. These are often presented to the plant as a series of corporate staff visits, all requiring the plants time and attention. If the plant acted on the sum of these "required programs" with full gusto, they would quit producing product. Their time and attention would be taken in team meetings, developing procedures, and trying to rationalize the differing demands of each program.

Fortunately, plant personnel know their jobs are tied to making budgeted quantities of quality product. So their view of change is to ignore these directives until they evaporate from lack of momentum and commitment. Their approach is well-rewarded; most programs do indeed get replaced by the next wave of "best practice" from corporate. I apologize to readers who have corporate development roles, as I may seem cynical. But in my experience the only change that lasts in the plant are those things that make sense to the guy on the plant floor. Safety, for instance, is in his best interest. It may have taken a decade or more of emphasis on safety, but everywhere we go we see good results and lasting change.

Some criteria for change of any kind to take hold in the plant are:

1. Intellectually it makes sense to the plant population. That sense means that improved productivity will *likely* result from such a program.
2. The plant population has a major say in how it will happen in their environment. They have the power, collectively, to determine *whether* it will proceed, and *how* it will proceed.
3. They see true commitment to the results. That means a number of things.
    - i. Some executive's future is tied to making this happen
    - ii. It has worked somewhere else that is similar enough to their environment
    - iii. The leadership team are all on board, no quibbling or sidebars
    - iv. Results are measured and posted visible locations in the plant
    - v. Valuable line people are assigned to the job, taken from other important tasks

2.4 Ownership.

How do we *enable* our work with our clients? What might apply to your own organization?

- **Leadership Consensus to Proceed.** SAMI won't proceed with billable work unless and until we know our client's leadership team, at the appropriate level, has consensus to proceed. Sometimes that means refusing an order for services from a plant manager, for instance, if we don't think the operations manager is fully supportive. To accept work under these conditions violates a cardinal rule, namely: *Anyone who has not been consulted does not feel he has to support the decision.* No matter how assured the person at the top of the organization is that the group will follow his decision, our experience is that lack of commitment by the *entire* leadership team is the number one cause of failure.
- **Worker Consensus to Proceed.** We usually begin our billable engagements with our clients with a baseline assessment of current processes, practices and results. Our clients typically believe this is because we need the data to know how to improve. The more important reason by far, however, is to engage the organization in a decision-making process that includes representation from all areas. Our assessments are designed to touch the greatest number of people practical, to solicit from them their issues, ideas and experience. There is an interesting pattern we almost always see. Leadership wants us to get the hourly workers to be willing to change; the hourly workers in turn challenge leadership to do its job and lead with strength of purpose, consistency and with high standards. The assessment process brings their views together, enabling them to see they want the same results: a productive, safe and competitive workplace where people are valued.
- **Develop a Workable Process and Passionate Owners.** Virtually every engagement we participate in has a work process design phase. The designers, typically a team of 8-10 part-time people, represent all types of jobs and all levels of the organization. We ask for leaders, even if those leaders can be negative at times. This team goes through the forming, storming, norming and performing stages of development, and we are careful to prepare them for the "J" curve effect (they go down emotionally before they go up). Their product is a completely thought out work management process, with all the details that will enable it to work in their environment. Our experience is that the product is 95% the same as virtually any other plant's. The 5% difference is critical, though, in practical workability. But the most important result of the design is a team of people who see the future and are passionate about making that future happen.
- **Client-Driven Implementation.** Only when workers see peers passionate about change will they pay attention. Outsiders (consultants) are seen a nuisances to be avoided. But if your respected peer is deeply committed to a new method of work, you will pay attention. And if he is willing to risk your relationship by making it not-optional, then you will believe it is a worthy change. Our consultants *support the client* in making the changes, not the other way around.

You should note that most of what we have written here is about enabling change, not about the core maintenance or production process. We assure that we have qualified experts in the process on the teams. Our clients seldom fail because they don't understand best practices, but because they can't get them implemented. We have learned to focus on change because it is the critical barrier for success!

2.5 Development.

It is possible that your people can work at much higher levels than they are today. In a reactive environment vs. a proactive environment these are frequent roles:

Job/Role	Reactive Environment	Proactive Environment
Craftsman	Component Replacer	Troubleshooter/RCF analyzer
Operator	Victim, Problem Identifier	Proactive Worker, Minor Maintenance
Supervisor	Expeditor	Work enabler, coordinator, troubleshooter
Engineer	Troubleshooter	Equipment defect elimination/optimization

Changing these roles is partially a matter of removing obstacles to being proactive and clarifying expectations, roles and responsibilities. But to a significant extent there is a requirement to assist people to be able to fill new roles. This is training, but much more than training. It involves coaching and testing the limits of the individuals in the job. Some operators are mechanically inclined, and some are not. Some will be eager to take on new roles, and some very resistant. Development takes time and energy for a supervisor to understand what is possible, and work with each of his people on a specific development program, customized to the specific task, and the native abilities of the worker. A training organization can be invaluable here, but the task cannot be delegated. The approach is rifle shots, not grandiose, one-size-fits-all programs. **Empowerment.** The "E" word has lots of bad connotations from the failures of quality programs in the 80's and early 90's. The popular method of empowerment was a week's worth of training in "soft skills", and an admonition that they should step up and be their own bosses. The result was lack of direction, anger, *disempowered* supervisors and management, and a decrease in productivity. Lee Solomon, founder of Soloman Associates who benchmark most of the world's refineries, once told me: ***"There is a high, negative correlation between implementation of self-managed work teams and performance".*** Empowerment as implemented not only didn't work, but it made things worse.

What ***is*** empowerment? It is enabling a worker to do more, and to take responsibility for his own performance. How can this best be done? First, by having a disciplined and well defined system to follow, and <u>enabling</u> the worker to be successful in the context of the system. Next, is to <u>develop</u> the worker to be successful in an expanded role. Finally, we give the worker the tools to understand whether he is mastering the job. This includes measures, feedback, coaching and encouragement. Empowerment is the **result** of a disciplined system of work, *not a prerequisite.*

3 CONCLUSION

Strategic Asset Management (SAM) is the systematic process that enables the dream of *Operations Excellence.* It emphasizes a logical approach to best practices through the developmental stages of the EXECUTE Triangles. Functional excellence will never be enough, however, to be the best. LEAD functions are the glue that brings all the pieces together in an optimized set of systems, especially through the mechanism of the *Managing System and Strategic Plan.* Finally, we can only be as successful as our workers endorsement and participation in these functional excellence practices. We must ENABLE our people to bring us the success we all desire.

We really become the best if we start our journey with the right model.

4 REFERENCES

[1] All thoughts, ideas, theories and strategies are the sole property of S. Bradley Peterson, President & CEO of Strategic Asset Management Inc.

5 APPENDIX:

Figures 1 thru 8

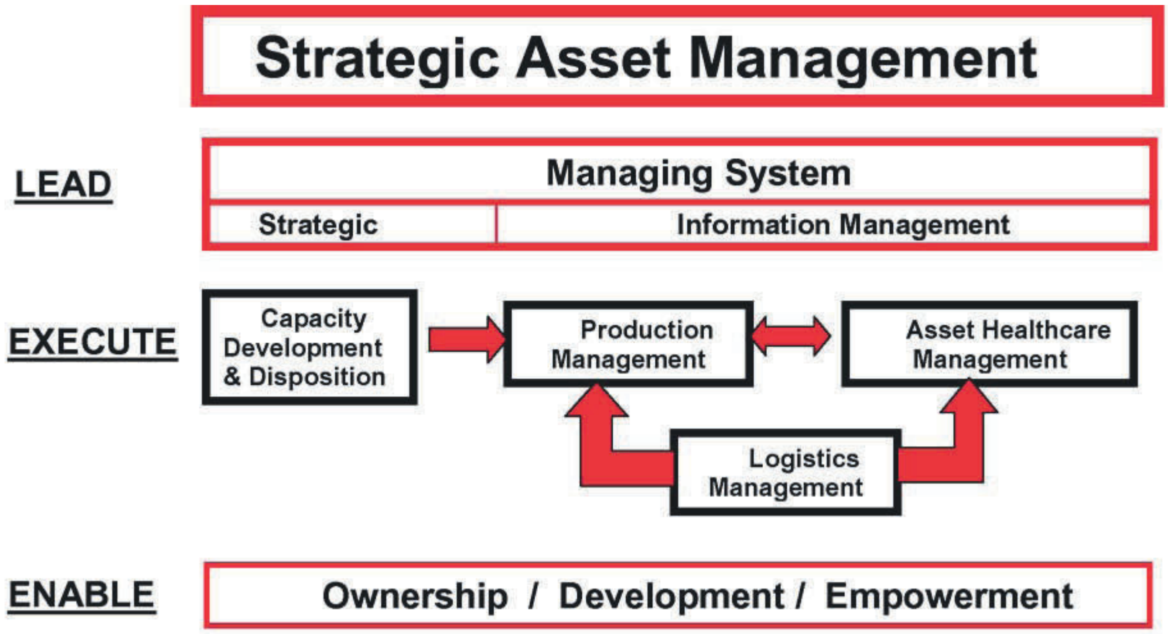


Figure 1: Strategic Asset Management Model (SAM Model)

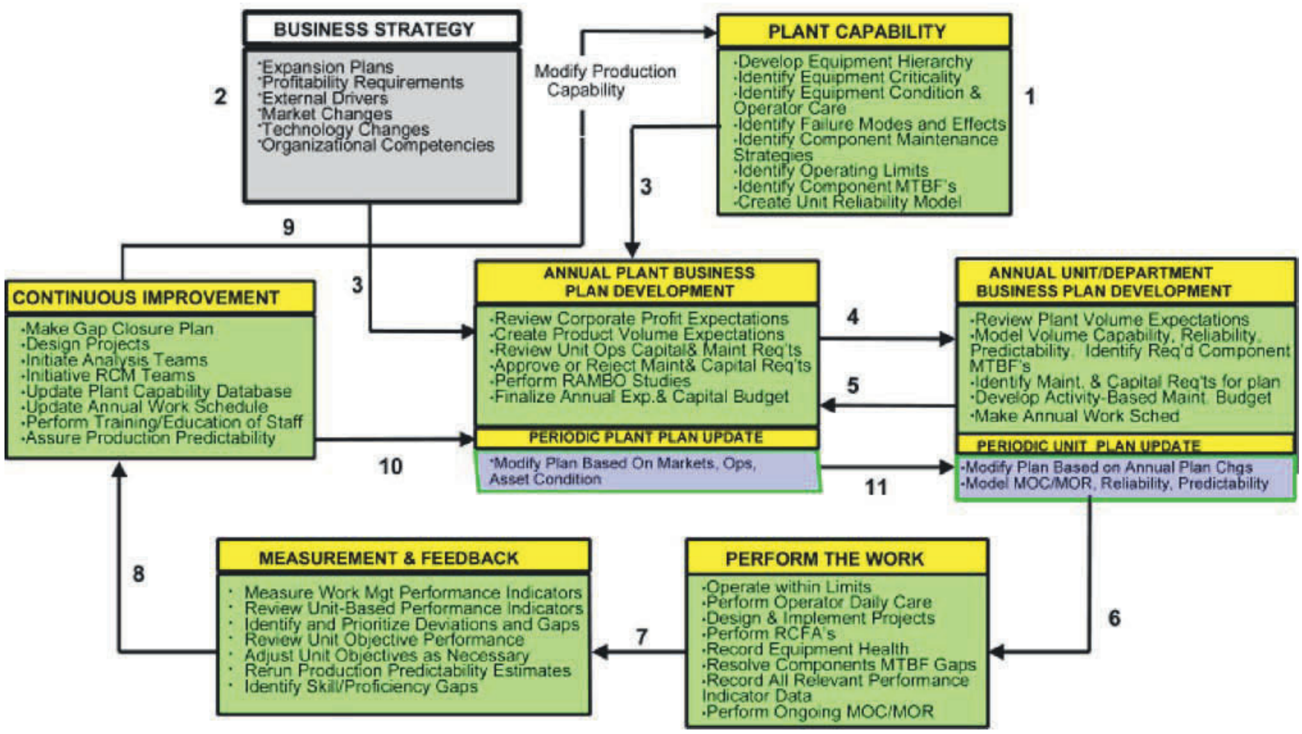


Figure 2: Managing System Model

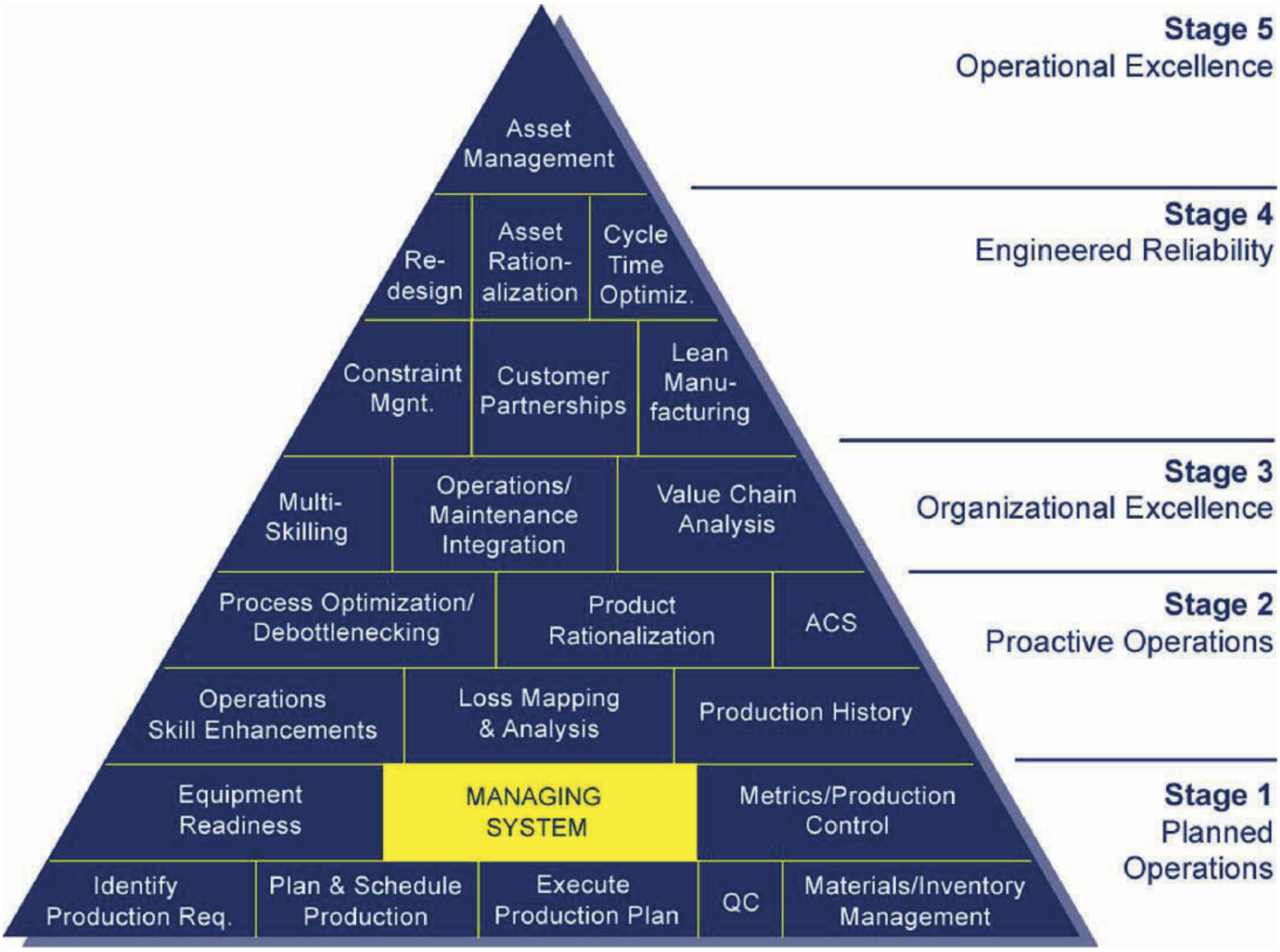


Figure 3: Production Triangle Model

Stage \ Class	Low Performing	Competent	High Performing
Stage 1 Planned Production	• Production Schedule Changes Daily • Little coordination between shifts • Frequent unplanned production interruptions • Raw materials have quality problems • Offspec product a way of life • No time for preventive maintenance • Frequent order expediting • Production targets seldom met	• Production schedule planned and executed • Raw materials available and meet spec • Equipment is made available for planned maint. • Operators capable normal production needs • Downtime and slowdown reasons logged • Equipment inspected, defect notifications done •Equipment permits is prepared for maintenance • Production targets usually met	• Minimum WIP inventory maintained • Scheduling is exact for materials, staff, time • Production meets or exceeds Q1 spec • Equipment is prepared daily by operator • Preventive maintenance is a religion • High level of cooperation among functions • Operators show equipment ownership • Production targets routinely exceeded
Stage 2 Proactive Production	• Operating spec's undocumented or ignored • Operators not trained on equipment/process • Production bottlenecks frequent • Little or improperly used automation • Operators have little process visibility • DCS antiquated, unstable, circumvented • Equipment not designed/maintained to perform required function	• Unplanned events evaluated, corrected • Operators educated in process & equipment • Production bottlenecks identified and worked • Operating specs identified & surveyed • DCS upgraded to perform additional functions • Quality built into process, not inspection • Raw material acquisition quality specified	• Asset healthcare program for all equipment • Predictive techniques minimize downtime • Failure analysis done for 80% of failures • Results of RCFA's implemented routinely • Condition monitoring based on cost & risk • Prevention is a cultural imperative • Trades work on designing out failure
Stage 3 Organizational Excellence	• Training unconnected to real work practice • "Team" implementation causes confusion • Supervisors lack authority, accountability • Unclear roles and responsibilities • Skills flexibility in contract, not implemented • Operators don't inspect or maintain equipment • No individual performance targets/goals	• Supervisors have clear roles & accountability • Operators inspect, lube, prep equipment • Task teams form, identify solutions, disband • Most work is directly by a priority system • Natural work teams effective at routine maint. • Service contract exists between maintenance & operations • Craft multi-skilling and flexibility implemented	• Work teams flexible, self-directed • Supervisors coach & advise • Continuous improvement embraced, effective • Clear priorities established for work • Reward/Recognition support best practices • Skills predominate over functions • All staff systems competent
Stage 4 Engineered Reliability	• RCM implementation creates confusion, results not implemented • Jobs changed on paper, not in reality • Stages 1&2 are ignored to "eliminate work" • Emphasis on analysis, not implementation • Vendor reduction brings lower service levels	• Reliability models run for critical systems component MTBF specifications defined • Equipment standards implemented • Critical equipment assessed via RCM • Projects get input from maintenance, ops • Equipment types, models rationalized • Purchasing buys by lifecycle cost, not price	• Concurrent engineering assures RAMBO • Reliability tied to financial results • Lifecycle costs are the basis for decisions • Vendor contracts pay for function reliability • Production targets set by reliability models • Equipment failures are a rare occurrence
Stage 5 Asset Management	• Management unclear re: goals & methods • Equipment condition not factored into goals • Equipment run parameters changed daily to respond to market pressures • Too many priorities prevent focus • Poor understanding of plant potential/liabilities	• Clear organizational alignment • Goals cascaded from plant level to individual • Production goals based on plant capability • Most work identified and planned prior year • Hourly help set unit goals & work improver's • Activity-Based Management implemented • Prod's reliability is part of product marketing	• Each employee knows & is rewarded for role • All decisions based on facts and models • 80% of work is preventive or project, & is identified prior to the start of the year • Production is 98% predictable • Lowest cost producer • Plant becomes corporate expansion site

RAMBO=Reliability, availability, maintainability, buildability, operability **WR=**Work request **RCFA=**Root Cause Failure Analysis **KPI=**Key Performance Indicator **OUE=**Overall Unit Effectives (OEE)

Figure 4: Production Maturity Matrix

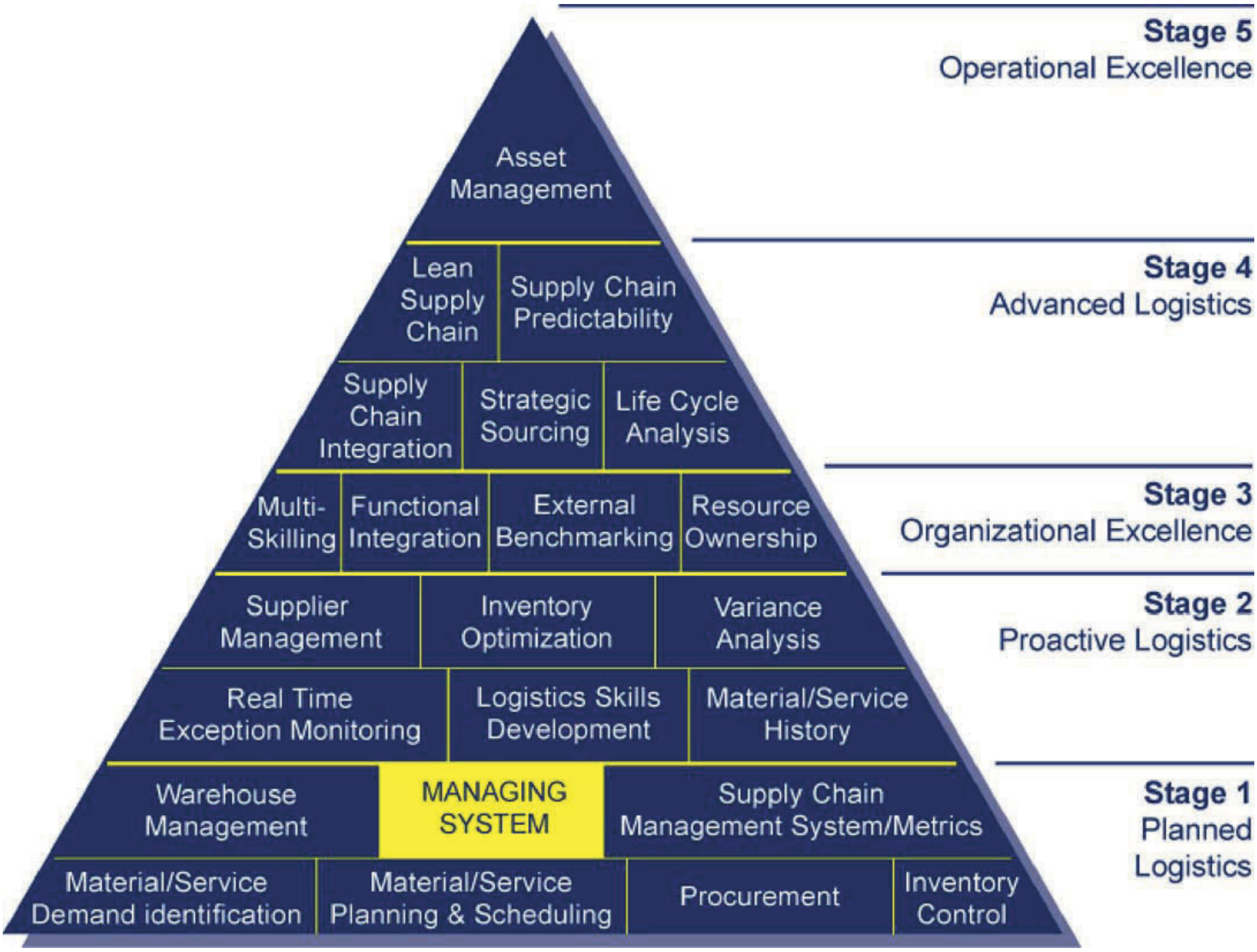


Figure 5: Logistics Triangle Model

Stage \ Class	Low Performing	Competent	High Performing
Stage 1 Planned Logistics	• Frequent reschedules of requirements •Excess of some items/frequent stock outs on others •Product/equipmentBOM's nonexistent or inaccurate • Unexecuted/inaccurate system transactions • Poor documentation/off-spec material a way of life • Frequent expediting • Delivery commitments often missed without notice • Inadequate security and control of inventory • Non-existent/inadequate performance indicators	• Most material requirements planned and stable • Controlled inventory levels with occasional stock outs •100% accurate BOM's for critical equipment/material • All material movement tracked in the system • Missed deliveries/incoming quality issues documented and resolved quickly • Supplier delivery dates usually satisfied on time/early • Some inventory counting routines in place • Occasional expediting • Performance indicators established and measured	• Requirements firm • Supplier delivery dates met >95% of the time • Early warming provided for missed deliveries • Storeroom improvement opportunities and implementation plans identified • Service improvement/inventory reduction opportunities and implementation plans identified • Formal Cycle Counting program in place to identify and resolve process issues; accuracy >95% • Expediting done proactively based on system data
Stage 2 Proactive Logistics	• Logistics personnel not trained in the process • No Supplier performance data available • Unavailable/inadequate supply chain planning and performance data • Supplier proliferation; suppliers selected randomly • Inventory managed strictly to assure service • Capabilities of the system not known/not understood • Planning data/reports unavailable or not being used	• Adequate job-related training, primarily OJT • Some data/reports available to track performance to plan • Supplier based controlled; suppliers selected primarily based on price and/or delivery • All materials purchased and stored using traditional methods • Inventory managed to balance services and investments • System capabilities understood but underutilized	• Required knowledge, skills and abilities identified and verified through formal training • Certification criteria in place to monitor and manage supplier base and performance; suppliers selected from qualified list based on criteria • Alternative procurement and stocking methods implemented to optimize service and costs • The system is capable of extracting required data and producing real-time exception notices
Stage 3 Organizational Excellence	• Unclear definition of roles and responsibilities • Focus on internal and individual performance • Nonexistent goals and objectives • Silo mindset • Resistance to change • No ownership processes	• Documented roles and responsibilities • Team-based work effective with focus on organizational goals and objectives • Acceptance of change • Individuals understand roles and responsibilities of personnel with other functions/from other departments • Multi-skilling implemented	• Cross-functional teams take ownership for resources/ processes and focus on plant/corporate goals and objectives as well as individual • Continuous improvement embraced, effective • Reward/recognition support best practices • Skills predominate over functions • Individuals can perform duties of personnel with other functions/from other departments
Stage 4 Advanced Logistics	• Suppliers are adversaries • Supply Chain/Logistics activities driven by reactive tactics • Numerous handoffs result in inefficiency and delays • Material specifications don't follow form, fit and functional needs • Equipment/process output unpredictable; customer requirements frequently missed without warning	• Suppliers are at arm's length • Supply Chain/Logistics activities driven by proactive planning • Material and transactions follow logical plan • Material specifications meet specific needs • Most customer requirements met on time or early	• Suppliers are strategic partners • Supply Chain/Logistic activities driven by strategic evaluation of total life cycle cost • Supply Chain extends to the Customer, and non-value-added functions have been eliminated • Material interchangeable in different applications without impacting quality, reliability or safety • Commitments met >95% of the time with early warning for potential missed deliveries
Stage 5 Asset Management	• Management unclear about goals and methods • Established processes circumvented in reaction to changing priorities • Commitments changed based on "squeaky wheel"	• Clear organizational alignment • Goals cascaded from plant/corporate level to individual • Commitments made based on supply chain capability and established plan • Logistics employees help establish goals and improvement plans • Supply Chain reliability is part of product marketing	• Each employee knows and is rewarded for his role • Established processes used to manage/align conflicting priorities • All decisions based on facts, models, and strategic business rules • Plant is lowest cost producer and becomes the benchmark for other sites

RAMBO=Reliability, availability, maintainability, buildability, operability **WR=**Work request **RCFA=**Root Cause Failure Analysis **KPI=**Key Performance Indicator **OUE=**Overall Unit Effectives (OEE)

Figure 6: Logistics Maturity Matrix Model

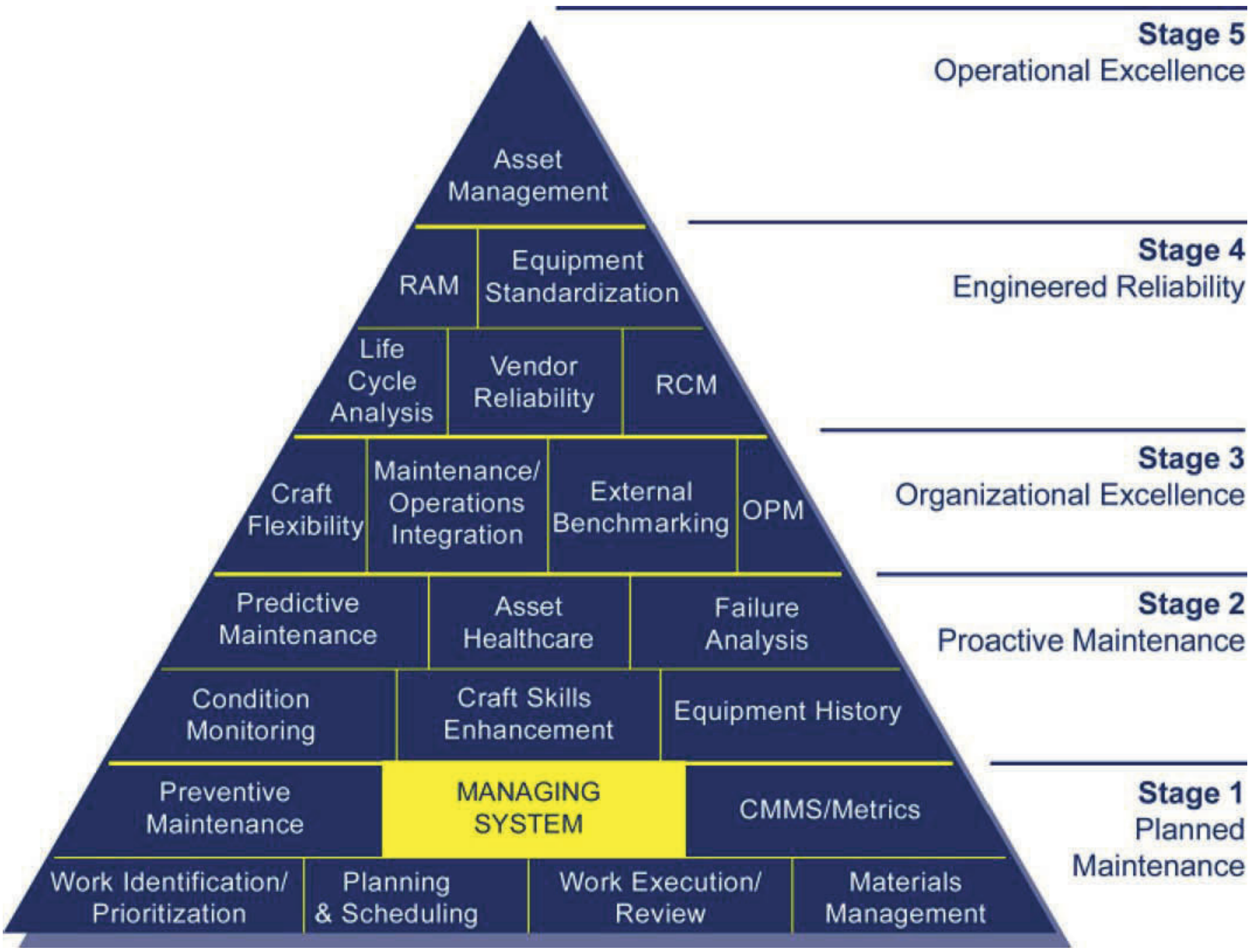


Figure 7: Asset Healthcare Triangle Model

Stage / Class	Low Performing	Competent	High Performing
Stage 1 Planned Maintenance	• "Fires" determine priorities • Breakdowns frequent • Maintenance equates to repair • Ineffective work orders, plans, controls • Stores service levels low • Poor operator/ maintenance relationships • Unable to meet production budgets	• Most work planned, scheduled • PM implemented, defects identified & corrected • Trades competent at most repairs • Computerized work system implemented • Stores service levels support operation • Operators prep for repairs • Maintenance goals & KPI's in place, reported	• All work prioritized, scheduled w/ prod's • PM hours and WO's exceed repairs • CMMS fully utilized • Parts kitted for jobs, 2x annual turns • Operators inspect, create WO's • Long-range scheduling implemented • Turnarounds well planned, executed
Stage 2 Proactive Maintenance	• Equipment history unusable • Failure analysis infrequent, ineffective • PM's don't match equipment or need • Inspections don't get done • Equipment criticality based on emotion • Condition monitoring sporadic, ignored • No comprehensive asset care program	• Equipment history is accurate and used • Critical assets have clear prevention plans • High percentage of PM compliance • Inspections yield WRs, scheduled & done • Condition monitoring widely used, trusted • Trades competent at most repairs • Failure analysis system established, effective	• Asset healthcare program for all equipment • Predictive techniques minimize downtime • Failure analysis done for 80% of failures • Results of RCFAs implemented routinely • Condition monitoring based on cost & risk • Prevention is a cultural imperative • Trades work on designing out failure
Stage 3 Organizational Excellence	• Training unconnected to real work practice • "Team" implementation causes confusion • Supervisors lack authority, accountability • Unclear roles and responsibilities • Skills flexibility in contract, not implemented • Operators don't inspect or maintain equipment • No individual performance targets/goals	• Supervisors have clear roles & accountability • Operators inspect, lube, prep equipment • Task teams form, identify solutions, disband • Most work is directly by a priority system • Natural work team effective at routine maintenance • Service contract exists between maint. & ops • Craft multi-skilling and flexibility implementation	• Work teams flexible, self-directed • Supervisors coach & advise • Continuous improvement embraced, effective • Clear priorities established for work • Reward/Recognition support best practices • Skills predominate over functions • All staff systems competent
Stage 4 Engineered Reliability	• RCM implementation creates confusion, results not implemented • Jobs changed on paper, not in reality • Stages 1&2 are ignored to "eliminate work" • Emphasis on analysis, not implementation • Vendor reduction brings lower service levels	• Reliability models run for critical systems • Component MTBF specifications defined • Equipment standards implemented • Critical equipment assessed via RCM • Projects get input from maintenance, ops • Equipment types, models rationalized • Purchasing buys by lifecycle cost, not price	• Concurrent engineering assures RAMBO • Reliability tied to financial results • Lifecycle costs are the basis for decisions • Vendor contracts pay for function reliability • Production targets set by reliability models • Equipment failures are a rare occurrence
Stage 5 Operational Excellence	• Management unclear re: goals & methods • Equipment condition not factored into goals • Equipment run parameters changed daily to respond to market pressures • Too many priorities prevent focus • Poor understanding of plant potential/ liabilities	• Clear organizational alignment • Goals cascaded from plant level to individual • Production goals based on plant capability • Most work identified and planned prior year • Hourly help set unit goals & work improvements • Activity-Based Management implemented • Production reliability is part of product marketing	• Each employee knows & is rewarded for role • All decisions based on facts and models • 80% of work is preventive or project, & is identified prior to the start of the year • Production is 98% predictable • Lowest cost producer • Plant becomes corporate expansion site

RAMBO=Reliability, availability, maintainability, buildability, operability **WR=**Work request **RCFA=**Root Cause Failure Analysis **KPI=**Key Performance Indicator **OUE=**Overall Unit Effectives (OEE)

Figure 8: Asset Healthcare Maturity Matrix Model

A DATA QUALITY FRAMEWORK FOR ENGINEERING ASSET MANAGEMENT

Shien Lin [a], Jing Gao [a] and Andy Koronios [a]

[a] *School of Computer and Information Science, University of South Australia, Mawson Lakes Boulevard, MAWSON LAKES SA 5095, Australia.*

Abstract: Data Quality (DQ) is a critical issue for effective engineering asset management (AM). DQ problems can result in severe negative consequences for an organisation. Several research studies have indicated that most organizations have DQ problems. This research aims to explore DQ issues associated with asset management in engineering organisations. The study first develops an AM specific DQ framework, and then tests it in a preliminary case study of two large Australian engineering organisations. The empirical findings of the DQ issues from the research are used to validate the proposed AM DQ framework. This study provides a better understanding of DQ issues for engineering asset management. This in turn will assist in providing useful advice for improving DQ in this area, leading to activities which will help ensure DQ. The research suggests that the importance of DQ issues for engineering asset management is often overlooked; thus, there is a need for more scrutinised studies in order to raise general awareness.

Key Words: Data Quality, Engineering Asset Management

1 INTRODUCTION

Industry has recently put a strong emphasis on to the area of asset management (AM). In order for organizations to generate revenue they need to utilize assets in an effective and efficient way. Often the success of an enterprise depends largely on its ability to utilize assets efficiently. Therefore, asset management has been regarded as an essential business process in many organizations. Furthermore, as companies today are running leaner than ever before, physical assets are being pushed to their limits, as engineering enterprises attempt to continuously drive more productivity out of their equipment, in order to improve their bottom lines. Consequently, physical asset management is moving to the forefront of contributing to an organization's financial objectives.

Nevertheless, there is strong evidence that most organisations have far more data than they possibly use; yet, at the same time, they do not have the data they really need (Levitan and Redman, 1998). Modern organizations, both public and private, are continually generating large volumes of data. On a personal level, according to Gartner Research (2005), each person on the planet generates an average of 250 Mbytes of data per annum, with this volume doubling each year. At the organizational level, there are incredibly large amounts of data, including structured and unstructured, enduring and temporal, content data, and an increasing amount of structural and discovery metadata. According to Hall (2005) of Computerworld, the situation will only get worse with the introduction of new data-generating technologies, such as Radio Frequency Identification (RFID), which will balloon the amount of data that has generated and accelerate this growth well beyond what IT organizations have experienced. And the trend of massive growth in data volumes will continue with no end in sight. Outside the business environment, there is an increasing number of embedded systems such as condition monitoring systems in ships, aircraft, process plants and other engineering assets, all producing gargantuan amounts of data. Despite this apparent explosion in the generation of data it appears that, at the management level, executives are not confident that they have enough correct, reliable, consistent and timely data upon which to make decisions. Many say “we are drowning in data and are starved of information”.

This lack of data visibility and control often leads to decisions being made more on the basis of judgment rather than being data driven (Koronios, Lin & Gao 2005). This can lead to less effective strategic business decisions, an inability to reengineer, mistrust between internal organizational units, increased costs, customer dissatisfaction, and loss of revenue. In some cases, it could also lead to catastrophic consequences such massive power failures, industrial or aviation disasters. Without good data, organisations are running blind and make any decision a gamble (ARC, 2004). They can’t make good decisions because they have no accurate understanding of what is happening within their company or the marketplace. They rely on intuition, which is dangerous in a fast-moving market with nimble competitors and finicky customers. Data and information are often used synonymously. In practice, managers differentiate information from data intuitively, and describe information as data that has been processed. Unless specified otherwise, this paper will use data interchangeably with information.

2 DATA QUALITY (DQ)

Numerous researchers have attempted to define data quality and to identify its dimensions (Kriebel, 1979; Ives, Olson, & Baroudi, 1983; Wang & Kon, 1993; Fox, Levitin & Redman, 1994; Wand & Wang, 1996; Wang & Strong, 1996; Shanks & Darke, 1998; Kahn, Strong & Wang, 2002). Traditionally, data quality has been described from the perspective of accuracy. However, this description has been challenged by a number of researchers (Strong, 1997; English, 1999; Salaun & Flores, 2001; Wang, Strong & Guauascio, 1994; Ballou, Wang, Pazer & Tayi, 1998; Orr, 1998), from the point of view that data quality should be defined beyond the accuracy dimension. Although there is no universal agreement on the meaning of "quality data", a common understanding found in the literature is that: "*quality data are data that are fit for use by the data consumer*" - Wang and Strong (1996). Orr (1998) also suggests that the issue of data quality is intertwined with how users actually use the data in the system, since the users are the ultimate judges of the quality of the data produced for them. With the aim of improving data quality, Wang (1998) suggests a Total Data Quality Management (TDQM) framework (define, measure, analyze and improve) for continuously managing data quality problems.

Dimensions of data quality typically include accuracy, reliability, importance, consistency, precision, timeliness, fineness, understandability, conciseness, and usefulness. Wand and Wang (1996) use ontological concepts to define data quality dimensions: completeness, unambiguousness, meaningfulness, and correctness. Wang and Strong (1996) categorize data quality into four dimensions: intrinsic, contextual, representational, and through accessibility. Shanks and Darke (1998) use semiotic theory to divide data quality into four levels: syntactic, semantic, pragmatic, and social. Recently, Kahn et al. (2002) have used product and service quality theory to categorize information quality into four categories: sound, useful, dependable, and usable.

Maintaining the quality of data is often acknowledged as problematic, but is also seen as critical to effective decision-making. Examples of the many factors that can impede data quality are identified within various elements of the data quality literature. These include: inadequate management structures for ensuring complete, timely and accurate reporting of data; inadequate rules, training, and procedural guidelines for those involved in data collection; fragmentation and inconsistencies among the services associated with data collection; and the requirement for new management methods which utilize accurate and relevant data to support the dynamic management environment.

Clearly, personnel management and organizational factors, as well as effective technological mechanisms, affect the ability to maintain data quality. Wang (1998) clarifies this relationship by drawing an analogy between manufacturing and the production of data. In this way they derive a hierarchy of responsibilities for data quality, ranging from management processes down to individual procedures and mechanisms. Their framework specifies a top management role for data quality policy, and a data quality management function to determine how that policy is to be implemented. This, in turn, should result in a data quality system for implementing data quality management, within which data quality control is enforced through operational techniques and activities. Data quality assurance then comprises all of the planned and systematic actions required to provide confidence that data meet the quality requirements.

3 ENGINEERING ASSET MANAGEMENT

Manufacturing assets are complex and expensive with multi-stage lifecycles. They begin as simple concepts to address an organization's needs and rapidly become physical entities that must be acquired, installed and handed-over to operating departments for use in generating revenues. During operation they must be carefully maintained to get maximum performance and longevity. Eventually they become obsolete and must be retired. Achieving maximum return-on-assets requires use of asset information and best practices for every activity, across all of these stages. In order to provide an in-depth understanding of the complex lifecycle asset management processes, the collaborative asset lifecycle management model (ARC Advisory Group, 2004) is adopted and illustrated in Figure 1.

There are three domains focused on the creation, use and management of the manufacturing assets in the model. The *Asset Lifecycle Domain* captures the processes related to asset creation, improvement and retirement. It includes key processes like design, manufacture, installation and decommissioning of complex facilities. The primary stakeholders in this domain are the asset owner and their Engineer, Procure and Construct (EPC) contractor. Once assets have been installed, they are operated and maintained by the respective groups in the owner/operator's organization. The *Asset Operation Domain* recognizes that operations consumes an asset's capabilities and these are restored periodically with parts and services acquired from the Original Equipment Manufacturers (OEM). The *Asset Performance Management Domain* includes those processes that occur during operation to monitor an asset's condition and manage the performance of the asset and the maintenance processes. Key players in this domain are the condition monitoring systems, Supervisory Control and Data Acquisition (SCADA) systems, the maintenance and technical staff and management. Asset management is at the heart of this model. This is a collaborative activity which includes the maintenance technicians, engineers and operators charged with the care and improvement of the manufacturing assets. And these stakeholders depend upon a reliable service network of service and parts providers.

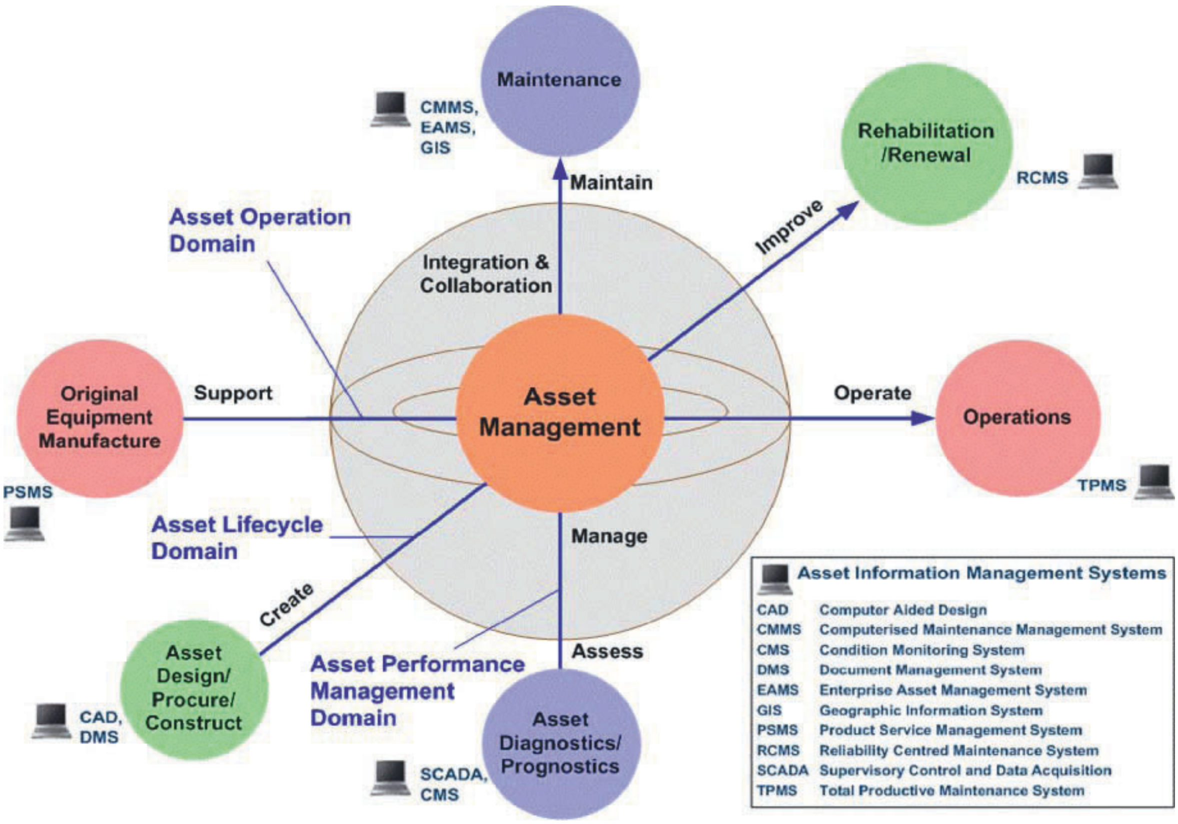


Figure 1: Collaborative asset lifecycle management and asset information management

(Source: Adopted and modified from ARC Advisory Group, 2004)

As discussed before, engineering assets have a unique set of characteristics as well as a sophisticated long process with different lifecycle stages. The process itself requires substantial information to be collected, processed, stored, analysed and further used throughout all stages of a typical asset's lifecycle. Because of the distinctive differences with regard to information systems and data requirements between general business environment and engineering asset management environment as discussed from previous research (Lin, Gao & Koronios, 2006), there is a need for an AM specific DQ framework to help obtain an insightful and overall understanding about what DQ issues are in AM, where the DQ problems are, why they have emerged, and where improvements can be made.

4 THE AM DQ FRAMEWORK

Based on the discussions on various data quality issues (Wang, Storey & Firth, 1995; Gelle, & Karhu, 2003; Xu et al., 2002; Xu et al., 2003; GAO, 2004) and the unique characteristics of asset management, an AM DQ framework was developed as shown in Figure 2. This framework is useful to guide the research into DQ issues in asset management, because it highlights the three root perspectives on data quality problems, illustrates how they emerge during the process of asset management; and outlines the basic data quality management criteria.

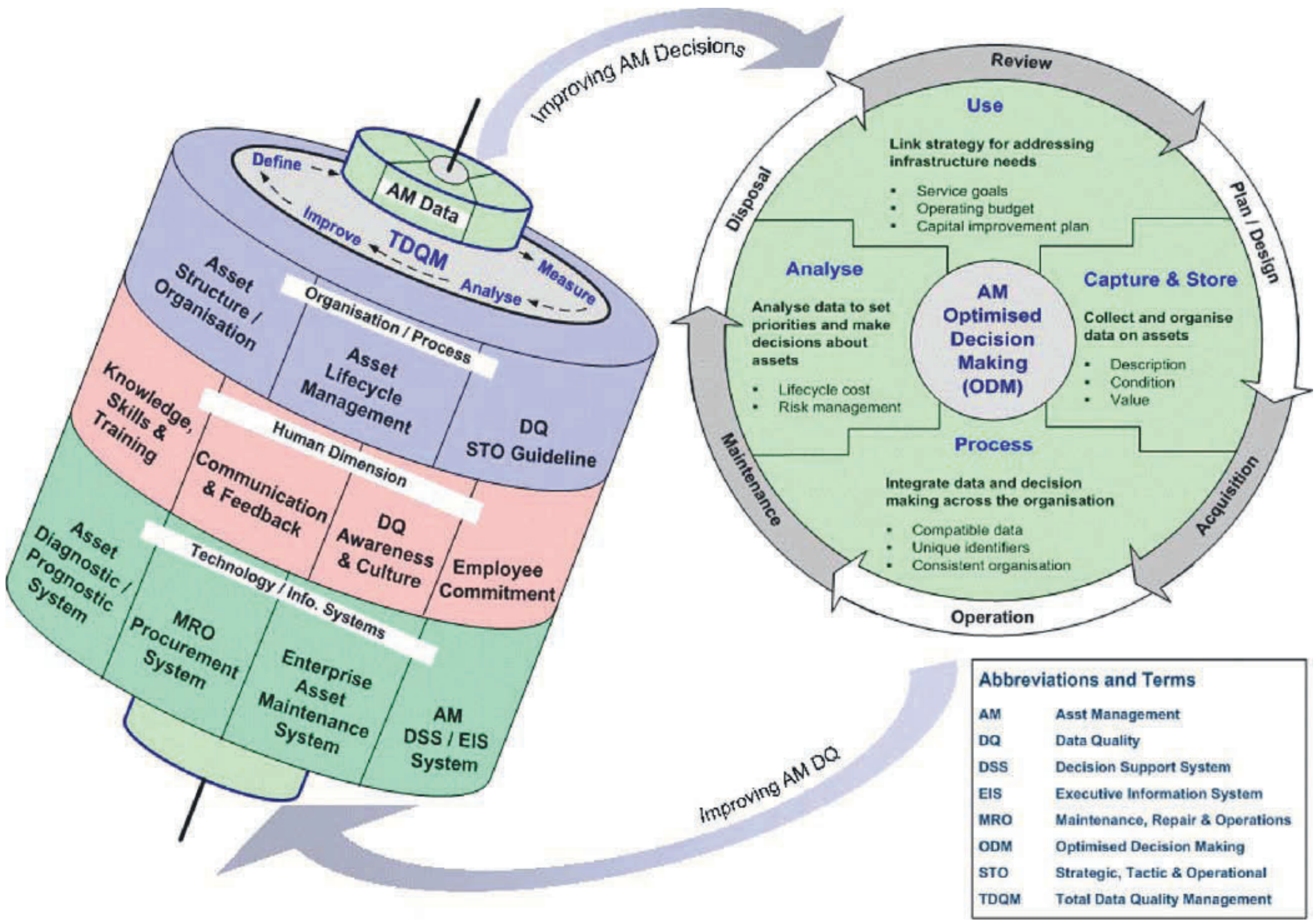


Figure 2: The AM DQ Framework (Source: Developed by the authors)

TOP Perspectives

Due to the uniqueness of engineering assets together with sophisticated long AM processes, engineering asset management demands total technological, organizational and personnel (TOP) supports. Mitroff and Linstone (1993) argue that any phenomenon, subsystem or system needs to be analysed from what they call a Multiple Perspective method – employing different ways of seeing, to seek perspectives on the problem. These different ways of seeing are demonstrated in the TOP model of Linstone (1999) and Mitroff and Linstone (1993). The TOP model allows analysts to look at the problem context from either Technical or Organizational or Personal points of view:

- The technical perspective (T) sees organizations as hierarchical structures or networks of interrelationships between individuals, groups, organizations and systems
- The organisational perspective (O) considers an organization's performance in terms of effectiveness and efficiencies. For example, leadership is one of the concerns.
- The personal perspective (P) focuses on the individual's concerns. For example, the issues of job description and job security are main concerns in this perspective.

Mitroff and Linstone (1993) suggest that these three perspectives can be applied as "three ways of seeing" any problems arising for or within a given phenomenon or system. Werhane (2002) further notes that the dynamic exchanges of ideas which emerge from using the TOP perspectives are essential, because they take into account "the fact that each of us individually, or as groups, organizations, or systems, creates and frames the world through a series of mental models, each of which, by itself, is incomplete". In other words, a single perspective on the problem context is not sufficient to elicit an insightful appreciation of it.

It is found that AM DQ requirements can be best described by using the TOP multiple-perspectives approach, which has close conformability with the key DQ research framework proposed by Wang, Storey & Firth (1995). Based on the previous discussion, it is felt that the process of modern asset management consists of the adoption of enterprise-wide information systems, various business processes, participants (e.g. maintenance people, managers, etc) and organizational policies (and business goals, structures, etc). It is thought that Linstone's TOP approach can be used to establish a building block for this framework (as a mean of identifying emerging DQ issues from AM).

4.1 Asset Lifecycle Process

An asset lifecycle starts at the time of designing the manufacturing system, and involves such activities as asset acquisition, operation, maintenance, decommissioning and replacement (Grobholz, 1988). Steed (1998) indicates that during its lifetime, the asset is subjected to a host of external factors: environmental conditions, system events, normal and abnormal loads, even changes brought about for whatever reason to the dielectric balance. At several critical stages, information is required on the condition of the assets. Knowing what to measure, how to measure it, and then what to do with the information becomes very important. Sandberg (1994) argues that contemporary asset management demands an elevated ability and knowledge to incessantly support asset management process, in terms of data acquisition, real-time monitoring, and computer supported categorization and recording of divergences from standard operations.

4.1.1 Technology

A variety of specialized technical, operational and administrative systems exist in asset management, which not only manage the operation of asset equipment but also provide maintenance support throughout the entire asset lifecycle. Most engineering organisations today purchase specialized systems from many suppliers. Some of these technical systems include asset diagnostic/prognostic systems, Supervisory Control and Data Acquisition (SCADA) systems, Computerized Maintenance Management Systems (CMMS), and Enterprise Asset Management systems. Normally these systems are bought from multiple vendors and each is specialized to accomplish its task. Unfortunately, this leads to an extremely difficult integration job for the end-user.

4.1.2 Organisation

The process of asset management is complicated. The lifecycle for a typical engineering asset involves several interdependent stages. Coordinating these processes as well as collaborating among other organizational processes affected by asset performance is vital to the effective management of engineering assets. Asset information is a key enabler in gaining control of assets. In order to facilitate AM data capture, process, storage, use and analysis, an asset hierarchy is essential. The objective of constructing an asset hierarchy is to provide a suitable framework for assets, which segments an asset base into appropriate classifications. The intent of the asset structure/organisation is to provide the business with the framework in which data are collected, information is reported, and decisions are made. In most cases, organizations work with an informal asset hierarchy. This often leads to data being collected to inappropriate levels, either creating situations where costs escalate with minimal increases in benefits, or insufficient information is available to make informed decisions (IPWEA, 2002). The information needs of the organization vary throughout the management structure. At the workface the key elements are operations, maintenance, and resource management, at a component level. At higher management levels this information needs to be aggregated to provide details on assets, facilities and systems as a whole in terms of finance, strategic and policy. The development of the asset identification needs to be appropriate for the asset hierarchy and complement the data/information needs of the organization (IPWEA, 2002).

4.1.3 People

In addition to the requirements for specialized IT/IS supporting systems, the AM process also require the involvement of assorted engineering and business stakeholders, internally and externally (Snitkin 2003). Because the impact of asset lifecycle management is too broad to be the sole responsibility of a single enterprise department, focusing only on the needs of AM personnel severely restricts the value that engineering asset management can bring to the organization. Supporting the needs of other stakeholders like management, finance & accounting, purchasing, product design, plant engineering, operations, sales, compliance, original equipment manufacturers (OEMs), and external service providers, has synergistic benefits that amplify the impact of all engineering asset management activities. As asset information is created, analysed and used throughout the whole asset lifecycle process by a variety of stakeholders, all stakeholders involved must be educated with regard to the impact of their actions and their roles, and take responsibility for supporting the collaborative asset management within the organization. Because of the diversity and high turnover of AM stakeholders, asset management outcome is also greatly associated with staff competency, communication & feedback, organizational culture, employee commitment.

4.2 Data Lifecycle Process

The sophisticated long AM process itself also requires substantial information support. Asset information is created throughout all stages of a typical asset's lifecycle, and underlies all the business processes in each domain. As data is a dynamic, fluid resource, it actually flows in a data lifecycle process. The major stages in the data lifecycle process include data acquisition, exchange and integration, storage and analysis. The DQ problems that may arise at each stage are different, and require different metrics as well as solutions [9]. Because of the continuum nature of data, DQ is not a one-time, fix-it-and-forget-it practice [13]. Building and keeping good quality corporate data takes constant vigilance and feedbacks in the context of the entire data lifecycle.

4.2.1 Data Acquisition

In increasingly rare cases, the data acquisition stage is preceded by a planning stage, where the amount of data gathered are planned and provided for. More often than not, however, it is the constraints on measuring devices and business exigencies that determine the kind of data that is collected, without the pre-planning stage. A major source of error at the data acquisition stage is the manual acquisition/entry of data. Manual entry can result in mis-typed data or lead to incomplete and missing data, non-standard entry, or duplicate entry. The lack of a pre-planning stage often introduces intractable DQ errors such as a mismatch between the scale of the problem and the scale of the hardware/software used. It can lead to short sighted or uninformed decisions that constrain the functionality of the data.

4.2.2 Data Integration

The most intractable DQ problems arise during data integration, the process by which multiple data feeds are brought together to create a rich and complete dataset. Most organizations are creating federated data, where data from different sources are stitched together to form a whole. The major DQ challenges during data integration are multiple data sources, different definitions, time synchronization, unconventional data, legacy systems, and sociological factors.

4.2.3 Data Storage

Once the data have been collected, they must be stored on devices in some format and with some storage management software. Physical storage is generally not a problem in these days. There are more concerns with the logical storage of the data – whether in a collection of files, in a database, or in some other structure. In any case, a schema must be developed, storage conventions decided upon, and metadata developed for the end user. The major DQ problems that arise during storage are: lack of awareness and planning, paucity of metadata, inappropriate data models, ad hoc modifications and changes, software and hardware constraints.

4.3 Total Data Quality Management (TDQM)

The process of engineering asset management is sophisticated, which requires gargantuan amounts of data to be collected from many different parts of the organization throughout all stages of a typical asset's lifecycle. This information must be maintained for many years in order to identify long-term trends. The AM engineering and planning processes use this information to plan and schedule asset maintenance, rehabilitation, and replacement activities. More asset data, however, does not necessarily mean better information, or more informed AM decisions. Many are finding it difficult to use this data. Most users are unable to translate the vast amounts of available asset data into meaningful management information to optimize their operation and control the total asset base. Researches found that although very large amounts of data is being generated from AM systems, little thought has been given to the quality of such generated data (Saunders 2004). Thus the quality of data from such systems may suffer from severe quality limitations. Miklovic (1998) reported that there are disconnects between the transaction-driven, product-centric business data environment and the continuous data, process-centric open control system and manufacturing data environments. The lack of process-to-product data transformation capabilities in linking business systems and plant floor AM applications have significant DQ consequences and thus negatively affect data-driven decision-making.

To achieve a state of high DQ, an organisation needs to take the approach of implementing Total Data Quality Management (TDQM) for understanding the DQ problems, identifying the causes, and developing the DQ improvement solutions. Wang (1998) adapted W. E. Deming's method of defining, measuring, analyzing, and improving products to the information manufacturing environment, and proposed TDQM, which emphasizes the importance of continuous improvement and delivery of high-quality information products. The defining component of the TDQM cycle identifies important IQ dimensions (Wang & Strong 1996) and corresponding DQ requirements. The measurement component produces DQ metrics. The analysis component identifies the root causes for DQ problems and calculates the impact of poor-quality information. Finally, the improvement component provides techniques for improving DQ. Different industries with different goals and environments can develop more specific and customized programs for DQ management to suit their own needs. Kovac, Lee & Pipino (1997) argue that regardless of differences organisations must develop a set of measures for their important DQ dimensions that can be linked to the organisational general goals and objectives. In the realms of improving data quality in engineering asset management, organisations need to focus on DQ control, DQ assurance and DQ management.

5 RESEARCH DESIGN

Based on the previous discussion about the EAM problems, an interview-based case study was designed to explore the DQ issues emerging within the chosen organizations. The organizations included two large Australian water utilities, as well as several of their subcontractors. A number of stakeholders at all levels of the organizations were interviewed, chosen on the basis of their experience in the use and management of engineering assets. The target organisations used a variety of information systems for EAM (e.g. GIS for asset location, Maximo for asset maintenance). Thirty interviews were conducted involving senior executives, asset managers, maintenance engineers, technicians, and data operators. In functional these stakeholders came from different position levels with different data roles at various office locations, including data provider, data custodian, data user, and data manager. In cases of conflicting issues, crosschecking for interviews was also conducted to validate the results.

Responses to our research questions were collated, stored, and analysed using qualitative data analysis software. This analysis allowed us to explore the raw data, identify and code the common themes, and identify relationships between themes in a rigorous manner. In analysing the collected data, an extensive examination of the viewpoints of various stakeholders was conducted. The views and actions of various interviewees in terms of their organizational interests were also examined. A very preliminary validation of the DQ for asset management model was achieved. While there may be some limitations in the approach used, we feel that the richness of the data collected far outweighed the methodological shortcomings of such an approach.

6 FINDINGS

The followings represent some of the preliminary findings based on using the TOP model. The integration of AM-related technical systems, as well as the integration between business systems and technical systems, is particularly important.

6.1 Integration of Technical Systems in Asset Management

A variety of specialized AM technical systems exist including reliability assessment systems, asset condition monitoring systems, asset maintenance systems. Such specialized systems are quite disparate, and acquired from multiple vendors, they often lead significant integration problems. There appears to be little cognizance when adopting business systems such as financial, human resource information systems of the need to ensure compatibility with technical systems such as asset register systems, condition monitoring systems. Most users are unable to translate the vast amounts of available asset data into meaningful management information to optimize their operation and control the total asset base. This has led to the notion of ‘islands of information’.

There are disconnects between the transaction-driven, product-centric business data environment and the continuous data, process-centric open control system and manufacturing data environments. Such disconnects make it extremely difficult to bring real-time information from the plant into business systems. The lack of process-to-product data transformation capabilities in linking business systems and plant floor EAM applications have significant DQ consequences and thus negatively affect data-driven decision-making.

6.2 Sensor Calibration and Integrity Check for Condition Monitoring

Interviews with asset maintenance field workers indicate that data captured by intelligent sensors may not always be accurate. Data capturing devices typically used in condition monitoring are electronic sensors or transducers, which convert numerous types of mechanical behavior into proportional electronic signals, usually voltage-sensitive signals, producing analog signals which in turn are processed in a number of ways using various electronic instruments. As signals are generally very weak, a charge amplifier is connected to the sensor or transducer to minimize noise interference and prevent signal loss. The amplified analogue signal can then be sent to filtering devices to remove or reduce noise, before being routed to a signal conditioner and/or analogue-to-digital converters for digital storage and analysis. To ensure the data received by the SCADA system conforms to the original signal data captured by sensors, integrity checks for signal transmission process and sensor calibration need to be performed and maintained. However, as the sensor calibration and integrity checks are often neglected in asset maintenance, the extent to which acquired data is correct and reliable was shown to be of concern with respondents.

6.3 Data Standard for Condition Monitoring Systems

Although it appears that condition monitoring equipment and systems are proliferating, an apparent lack of dialogue among vendors has led to incompatibilities among hardware, software and instrumentation. Data collected by current outdated equipment could become obsolete and inaccessible to new upgraded systems. To fully realize the integration of systems over the various levels of asset maintenance and management, new standards and protocols are needed. A focus on standardization of condition monitoring data modelling and exchange tools and methodologies, such as Standard for the Exchange of Product model data (STEP) is critical.

6.4 Database Synchronization

The capability of EAM systems can be enhanced through a link with GIS to provide the ability to access, use, display, and manage spatial data. The ability to effectively use spatial asset data is important for utilities with geographically dispersed utility networks. However, it was found that one of the most critical activities is to establish synchronization between the two database environments. One asset manager indicated that there has been an issue existed for overcoming the synchronization of asset register in a very common work management system with GIS in the company. Both automated and manual processes needed to be defined and implemented to maintain synchronization between the GIS and EAM databases. Database triggers and stored procedures need to be defined to automate the attribute update process maintaining synchronization between the GIS and EAM databases. Workflows and business rules must be developed for GIS and EAM data editing, to ensure synchronization from both applications.

6.5 Business Process Reengineering

Organisational fit and adaptation are important to implementation of modern large-scale enterprise systems. Like enterprise resource planning systems, EAM systems are also built with a pre-determined business process methodology that requires a fairly rigid business structure in order for it to work successfully. They are only as effective as the processes in which they operate. Companies that place faith in EAM systems often do so without reengineering their processes to fit the system requirements. Consequently, this often results in negative impacts on the effectiveness of both the EAM system and the AM practices. It was found that a mismatch existed between the business processes requirements of the EAM and actual practice in the organisation.

6.6 Data Recording

Research in data collection has found that DQ and validation effectiveness improve, the sooner the collected data is entered and the nearer the data entry is to the asset and its work. If a data entry point is remote from the asset, then the capability for accurately confirming the data is considerably reduced and the temptation to enter something - anything that the system will accept - is great. One manager said in the interview that “I feel that most of the (data) errors over time have been because of the lag between the field data and being continued in the computer somewhere….they (field staff) might wait a week before they complete their work order (entry)”. It was found that the longer the time lag between using the entered data and the time it was initially created, the less chance of cleaning up the data to make it useful.

6.7 Training

AM requires all aspects of training as well as appropriate documentation of the system. It was found that organisations tended to focus more on the “hardware” part of the systems development process, putting less effort on the “soft” part, that is, the training of how to operate and manage the system. Several respondents indicated that the training was tailored for the specific areas but the same for everyone and thus of limited use to some. Many respondents contended that what they knew about the system was in fact ‘self taught’. Awareness of issues such as how the data being collected was going to be used was not existent; yet they agreed that such knowledge would increase motivation and performance by the asset operators/technicians. Lack of training can have an adverse impact on information quality. It is easy for organisations to find reasons/excuses for avoiding adequate training for the staff and management.

6.8 Communication & Management Feedback

Competitive asset-intensive companies have reported that most of their asset improvements come from their workforce. Despite the fact that “people are our greatest asset”, evidence of the opposite is often found. People’s problems, relationships, aspirations and their personal agendas are seldom given any consideration. In the implementation and management of EAM systems it appears that this is not different and was quite evident in the responses. Respondents were quite convinced that the system implementation neglected the human dimensions and thus contributed to the partial failure of these systems.

“year after year they (field workers) filled out field data without feedback……if we did nothing, nothing happens so why bother?”

7 CONCLUSION

Several researches have indicated that maintaining the quality of data is often acknowledged as problematic, but is also seen as critical to effective decision-making in engineering asset management. Managing data quality is essential to any asset management programme. This research develops an AM DQ framework in order to help organisations and practitioners understand DQ problems, identify causes, and develop solutions in the realms of engineering asset management. The framework proposed in this paper provides a useful tool for planning the establishment of an awareness of DQ issues in managing assets. With the adoption of the DQ framework, researchers can obtain an insightful and overall understanding about what DQ issues are in engineering asset management, where the DQ problems are, why they have emerged, and where improvements can be made. Future research will further develop the framework and identify DQ toolkit (e.g. guidelines, functional requirements for asset data cleansing software) for the improvement of data quality associated with engineering asset management.

8 REFERENCES

[1] Levitan, A.V. and Redman, T.C., (1998) Data as a resource: properties, implications and prescriptions, *Sloan Management Review*, 40(1), 89-101.

[2] Steenstrup, K., (2005) Enterprise Asset Management Thrives on Data Consistency, Research Article, *Gartner Research*.

[3] Hall, M., (2005) Databases Can't Handle RFID, News Story, *Computerworld*.

[4] Koronios, A., Lin, S., & Gao, J., (2005) A Data Quality Model for Asset Management in Engineering Organisations, *Proceedings of the 10th International Conference on Information Quality (ICIQ 2005),* Massachusetts Institute of Technology, Cambridge, USA.

[5] ARC Advisory Group, (2004) Asset Information Management – A CALM Prerequisite, *White Paper.*

[6] Kriebel, C.H., (1979) Evaluating the quality of information systems, in *Design, and Implementation of Computer Based Information Systems,* Szysperski, N., and Grochla, E., eds. Germantown: Sijthroff and Noordhoff.

[7] Ives, B., Olson, M., and Baroudi, J.J., (1983) The measurement of user information satisfaction, *Communications of the ACM,* 26(10), 785-793.

[8] Wang, R.Y., and Kon, H.B., (1993) Toward Total Data Quality Management (TDQM), in *Information Technology in Action: Trends and Perspectives*, Wang, R.Y. (Eds), Prentice Hall, Englewood Cliffs, NJ.

[9] Fox, C., Levitin, A.V., and Redman, T.C., (1994) The notion of data and its quality dimensions, *Information Processing & Management,* 30(1), 9-19.

[10] Wand, Y., and Wang, R.Y., (1996) Anchoring Data Quality Dimensions in Ontological Foundations, *Communications of the ACM,* 39(11), 86-95.

[11] Shanks, G., and Darke, P., (1998) Understanding data quality in a data warehouse, *Australian Computer Journal*, 30(4), 122-128.

[12] Kahn, B.K., Strong, D.M., and Wang, R.Y., (2002) Information Quality Benchmarks: Product and Services Performance, *Communications of the ACM,* 45(4), 184-192.

[13] Strong, D.M., (1997) IT Prosess designs for Improving Information Quality and reducing Exception Handling: A Simulation Experiment, *Information and Management,* (31), 251-263.

[14] English, L.P., (1999) *Improving Data Warehouse and Business Information Quality: Methods for reducing costs and increasing Profits*, Willey & Sons.

[15] Salaun, Y., and Flores, K., (2001) Information Quality: Meeting the Needs of the Consumer, *International Journal of Information Management*, 21(1), 21-37.

[16] Wang, R. Y., Strong, D., and Guauascio, L. M., (1994) Beyond Accuracy: What data Quality Means to Data Consumers, *Technical Report TDQM-94-10*, *Total Data Quality Management Research Program*, MIT Sloan School of Management, Cambridge, MA.

[17] Ballou D., Wang R., Pazer, H., Tayi, G.K., (1998) Modeling Information Manufacturing Systems to Determine Information Product Quality, *Management Science*, 44(4), 462-484.

[18] Orr, K., (1998) Data Quality and System Theory, *Communications of the ACM*, 41(2), 66-71.

[19] Wang, R.Y., and Strong, D.M., (1996) Beyond Accuracy: What Data Quality Means to Data Consumers, *Journal of Management Information Systems,* 12(4), 5-33.

[20] [20] Wang, R., (1998) A product perspective on data quality management, *Communications of the ACM*, 41(2), 58-65.

[21] Segev, A., (1996) On Information Quality and the WWW Impact a Position Paper, *Proceedings of Conference of Information Quality.*

[22] Firth, C., (1996) Data Quality in Practice: Experience from the Frontline, *Proceedings of Conference of Information Quality.*

[23] Saraph, (1989) An Instrument for Measuring the Critical Factors of Quality Management, *Decision Sciences*, 20(4), 810-829.

[24] IPWEA, (2002) *International Infrastructure Management Manual*, Australia/New Zealand Edition.

[25] Eerens, E., (2003) *Business Driven Asset Management for Industrial & Infrastructure Assets*, Le Clochard, Australia,

[26] Woodhouse, J., (2001) Asset Management, The Woodhouse Partnership Ltd, 2001, online accessed on December 10, 2005 at http://www.plant-maintenance.com/articles/AMbasicintro.pdf.

[27] Woodhouse, J., (2003) Asset Management: concepts & practices, The Woodhouse Partnership Ltd.

[28] Koronios, A., and Haider, A., (2003) Managing Engineering Assets: A Knowledge Based Asset Management Methodology Through Information Quality", *E-Business and Organisations in the 21th Century,* 443-452.

[29] Snitkin, S., (2003) Collaborative Asset Lifecycle Management Vision and Strategies. *ARC Advisory Group*.

[30] Mitroff, I. I. and H. A. Linstone, (1993) *The Unbounded Mind: Breaking the Chains of Traditional Business Thinking,* New York, OxFord University Press.

[31] Linstone, H. A., (1999) *Decision Making for Technology Executives: Using Multiple Perspectives to Improve Performance,* Artech House Publisher.

[32] Werhane, P. H., (2002) Moral Imagination and Systems Thinking, *Journal of Business Ethics*, (38), 33-42.

[33] Wang, R.Y., Storey, V.C., and Firth, C.P., (1995) A Framework for Analysis of Quality Research, *IEEE Transactions on Knowledge and Data Engineering,* 7(4), 623-640.

[34] Gelle, E. & Karhu, K., (2003) Information quality for strategic technology planning, *Industrial Management & Data Systems,* 103(8), 633-643.

[35] Xu, H., Nord, J. H., Brown, N. & Nord, G. D., (2002) Data quality issues in implementing an ERP, *Industrial Management & Data Systems,* 102(1), 47-58.

[36] Xu, H., Nord, J. H., Nord, G. D. & Lin, B., (2003) Key issues of accounting information quality management: Australian case studies, *Industrial Management & Data Systems*, 103(7), 461-470.

[37] United States General Accounting Office (GAO), (2004) Water Infrastructure: Water Utility Asset Management, *GAO-04-461*, Washington, D.C.

[38] Miklovic, D., (1998) Plant Floor-to-Enterprise Application Data Transformation, *Research Note*, Gartner Group.

DISTRIBUTED ALGORITHM FOR LOCALIZATION OF LARGE SCALE SENSOR NETWORK

Kenji OGUNI [a], Hiroaki HONDA [a], Kok-How KHOR [a] and Junya INOUE [b]

[a] *Earthquake Research Institute, The University of Tokyo, Yayoi 1-1-1, Bunkyo, Tokyo 113-0032, Japan.*

[b] *The University of Tokyo, Hongo 7-3-1, Bunkyo, Tokyo 113-0033, Japan.*

Abstract: Localization (determination of the position of each sensor node) of a large scale sensor network is an important issue for applying network sensing technique to the health monitoring of the infrastructures. This paper presents an implementation of acoustic ranging and proposes a distributed algorithm for localization of sensor network. Relative positions between sensor nodes are estimated based on acoustic ranging through the inverse Delaunay algorithm. This algorithm localizes all the nodes simultaneously, thus, the accumulation of the error in the localization is suppressed. The robustness and the scalability of this algorithm are examined through numerical simulations. Noise tolerant acoustic ranging algorithm that employs digital signal processing techniques is implemented in an off-the-shelf sensor platform (Mica2). Experiments show that this acoustic ranging algorithm is sufficient to give average range estimation error below 10 cm. Together with this acoustic ranging algorithm, the inverse Delaunay algorithm is also implemented in Mica2. This enables automatic localization of sensor network. Field experiment was conducted to evaluate the accuracy of the localization of the system.

Key Words: Robust localization, Sensor network, Acoustic Ranging, Delaunay

1 INTRODUCTION

Recent developments in sensor and wireless communication technologies make it possible to distribute sensor platform in the environment[1]. One of the final targets to apply this network sensing system is civil infrastructures. High resolution sensing and on-site simulation using dense sensor network on civil infrastructures can be regarded as a typical example of sensor embedded society. The sensor network applied to civil infrastructures should cover wide area with high spatial resolution. This results in the requirement for numerous sensor nodes with low cost. This cost consists of cost for sensor itself and the installation cost. Especially to reduce the installation cost, automatic localization of the sensor nodes is needed. Although the required accuracy of localization depends on the application, we aim at the accuracy of less than 10cm for all the sensor nodes distributed over the space of the size of a whole city. Installation of GPS receivers on all the sensor nodes is not a realistic option because of the cost of available GPS receivers with the accuracy under consideration[2]. Motivated by this background, this paper presents an algorithm and an implementation of the sensor network system for robust localization. Localization of the sensor nodes in the network is achieved by a distributed algorithm called “Inverse Delaunay Algorithm.”

Localization of the sensor nodes presented in this paper uses relative distance between neighboring nodes measured by the acoustic ranging as input data. Direct application of the techniques in surveying (i.e., starting from reference nodes, localization is done on the nodes consecutively through the triangulation) does not work in this case because of the noisy measurement on the relative distance and the accumulation of the error. It seems that the distributed algorithm consisting of i) construction of local cluster of nodes with relative positions identified in the local coordinates for the cluster and ii) inter-cluster patching algorithm, works better for localization based on noisy measurement on the relative distance between nodes. Existing work based on this approach is summarized in Moore *et al.* [3] as attempts for graph realization problem. In this problem, reduction of the flexibility of the local cluster of the nodes (e.g., flip ambiguity, discontinuous flex ambiguity) is the major concern. The inverse Delaunay algorithm presented in this paper is categorized in the attempts for graph realization problem.

2 RELATIVE LOCALIZATION BASED ON INVERSE DELAUNAY ALGORITHM

Each and every sensor node in the randomly distributed sensor network is localized by the algorithm discussed in this section. Localization is achieved by clustering the neighbouring set of nodes with Delaunay polygons using the observed relative distance between sensor nodes as input data. This is opposite from Delaunay tessellation of a space for given arrangement of points. In this sense, this localization algorithm is called inverse Delaunay algorithm. Before discussing the details of the inverse Delaunay algorithm, mathematical definition and characteristics of the Delaunay tessellation/polygons are given in this section[4].

For a set of finite number of points $X = \{\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^n\}$ on 2D real space ($\Re^2$) with Euclid distance $d(\mathbf{x}, \mathbf{y})$ between points $\mathbf{x}$ and $\mathbf{y}$, define

$$V(\mathbf{x}^i) = \{\mathbf{x} \mid \mathbf{x} \in \Re^2, d(\mathbf{x}, \mathbf{x}^i) < d(\mathbf{x}, \mathbf{x}^j), i \neq j\}. \tag{1}$$

Then, $V(\mathbf{x}^i)$ becomes a convex polygon containing a point $\mathbf{x}^i$ and a set $\{V(\mathbf{x}^1), V(\mathbf{x}^2), \cdots, V(\mathbf{x}^n)\}$ gives a tessellation on $\Re^2$. This tessellation is called Voronoi tessellation for a set X and denoted as $V(X)$. Points in a set X are called mother points of $V(X)$, and $V(\mathbf{x}^i)$ is called a Voronoi cell for a mother point $\mathbf{x}^i$.

When $V(\mathbf{x}^i)$ and $V(\mathbf{x}^j)$ share a Voronoi edge (i.e., edge of the Voronoi polygons), conjugate geometry to Voronoi diagram can be obtained by connecting $\mathbf{x}^i$ and $\mathbf{x}^j$. This diagram is called Delaunay diagram and the tessellation using Delaunay diagram is called Delaunay tessellation. The vertices and the circumcenters of the Delaunay polygon correspond to the mother points in Voronoi tessellation and vertices of the Voronoi cells, respectively (see Fig. 1). The following characteristic holds for a Delaunay diagram in 2D space:

Polygons in the Delaunay diagram on $\Re^2$ are triangles with no mother point of $V(X)$ in their circumscribed circle

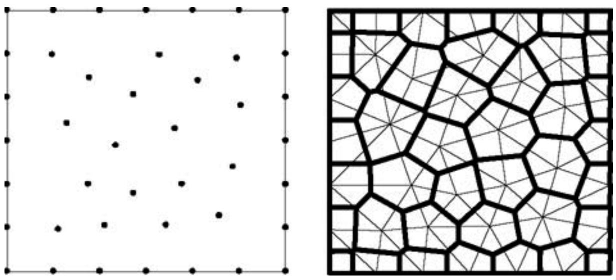


Figure 1 left: mother points (corresponding to the sensor nodes to be localized), right: Voronoi (thick) and Delaunay (thin)

The inverse Delaunay algorithm discussed in the following subsections consists of two steps, namely, 1) construction of local clusters (Delaunay clusters), and 2) merging procedure of local clusters. The characteristic of Delaunay diagram shown above is extensively used to increase the robustness of the first step (i.e., toughness of the local clusters against the measurement error).

Note that the whole discussion in this section can be easily applied to 3D real space by extending Eqn. (1) and the corresponding Euclid distance to those for $\Re^3$. In that case, a Voronoi cell, $V(\mathbf{x}^i)$ becomes a polyhedron and the Delaunay tessellation becomes a tessellation consisting of tetrahedrons. However, for the sake of simplicity, description of the algorithm is given for 2D problem setting henceforth.

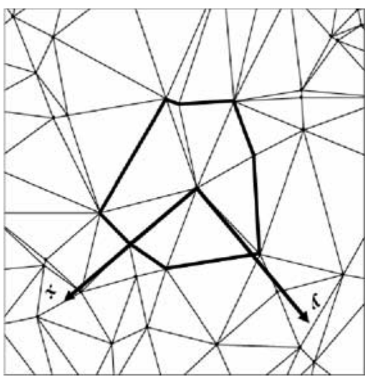


Figure 2 Typical example of the local cluster

2.1 Construction of local Delaunay cluster

In the first step of the inverse Delaunay algorithm, a local cluster is made for each sensor node (which results in a mother point in the Voronoi diagram made by localized sensor nodes). Figure 2 shows a typical example of a local cluster. The area surrounded by thick solid lines is a local cluster. In this local cluster, the center node is surrounded by the Delaunay triangles hence this cluster is called Delaunay cluster. All the nodes are localized in the local Euclidean coordinate system with the center node set as the origin. Also, the nodes surrounding the center node (satellite nodes) are numbered counter-clockwise in the local coordinate system. This numbering determines the direction of the face of the local cluster.

Each local cluster is generated in the manner as shown in Fig. 3.

1. A node is picked up as a center node (filled circle in the figure) and the closest neighboring node is identified based on the measurement of the relative distance. (Fig. 3 (a))
2. Form a triangle using above-mentioned nodes and another neighboring node. Examine this triangle to see whether it is Delaunay or not. Triangle with dotted line in Fig. 3 (b) is not Delaunay since other node is located in the circumscribed circle.
3. Local cluster is expanded by adding another Delaunay triangle which has a center node as one of the vertices (Fig. 3 (c)).
4. After surrounding the center node by Delaunay triangles, the satellite nodes are numbered. The Delaunay cluster shown in Fig. 3 (d) has the clockwise numbering. This means the local coordinate for this Delaunay cluster is flipped with respect to the global coordinate (which can be identified after the global localization is done).

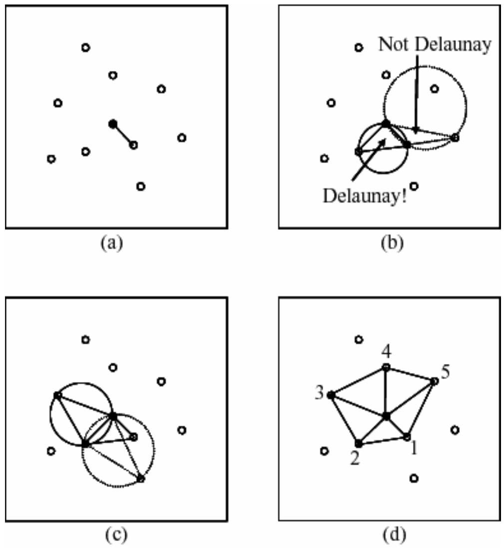


Figure 3 Construction of the Delaunay cluster

Note that Fig. 3 shows a conceptual view of the arrangement of the nodes and the local cluster. In reality, position of each node has not been identified before localization is done. All we have is the relative distance between nodes. Delaunay examination and expansion of local cluster solely depend on relative distance.

2.2 Merging procedure of local Delaunay cluster

Once the local Delaunay clusters are identified for all the nodes, those clusters should be merged together. Merging is achieved by identifying the necessary translation, rotation and flip for each cluster to satisfy consistency among local clusters with common set of nodes. The details of the merging procedure are as follows:

1. Clusters are divided into two groups, i.e., the atomic clusters and the bridging clusters. The atomic clusters are those with satellite nodes not being occupied by other clusters (i.e., polygons depicted by the edges with thick lines in Fig. 4). The bridging clusters are others. This categorization is done by first-come-first-serve basis in the current version of our source code. However, the categorization based on the reliability or the shape of the Delaunay cluster can be implemented.
2. Connect the atomic clusters (or groups) by the bridging clusters. If the order of numbering of the corresponding satellite nodes of both clusters (or groups) coincides, direction of the faces of these clusters (or groups) are different. Either one of them should be flipped. In Fig. 4, corresponding nodes in the atomic and the bridging cluster are numbered as $1 \rightarrow 2$ and $2 \rightarrow 3$, respectively, i.e., both have the ascending order. Therefore, the bridging cluster is flipped.
3. Apply the necessary translation, rotation and flip to the cluster or group to be consolidated.
4. Continue the second and the third process until all the clusters are merged into a group.

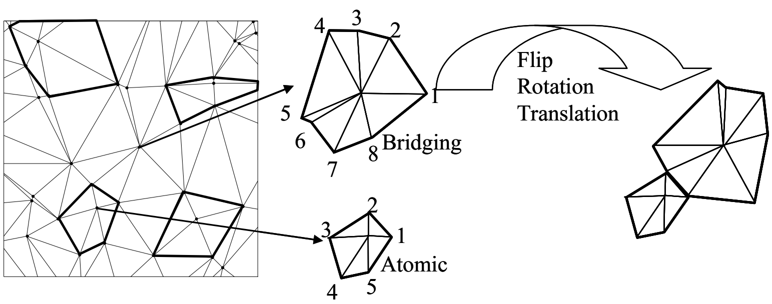
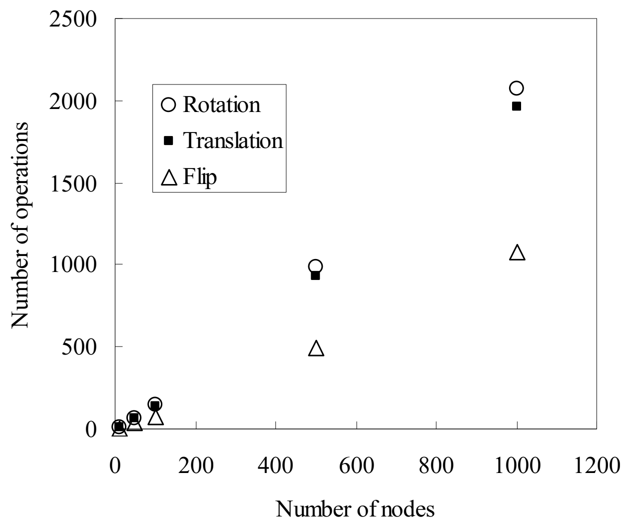


Figure 4 Cluster merging procedure

2.3 Computational complexity and robustness of the inverse Delaunay algorithm



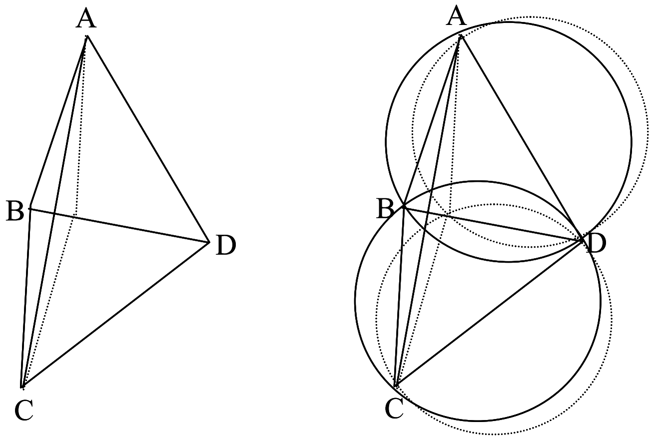


Figure 6 Robustness of the Delaunay cluster
left: the quadrilateral is not robust because of the measurement noise in the length BD
right: regardless of the position of node B, Delaunay is robust

Localization of sensor nodes based on relative distance between nodes is categorized in graph representation problem in computational geometry. As shown in the previous subsections, the proposed localization algorithm gives the relative location of the nodes using local clusters consisting of the Delaunay triangles. Thus, the algorithm is named as the inverse Delaunay algorithm.

One of the major advantages of the inverse Dealunay algorithm is its small computational complexity. Obviously, the construction of the local Delaunay clusters has the complexity of $O(n)$ where n is the number of the nodes. Figure 5 shows the total number of operation (i.e., translation, rotation and flip) in the merging procedure with respect to number of nodes. Figure 5 proves that computational complexity is $O(n)$ for merging procedure at least upto n=1000. This suppression of the complexity could due to the sparse arrangement of the atomic and bridging clusters.

Another major advantage of the inverse Delaunay algorithm is its robustness against the measurement error in ranging. To close the discussion on this algorithm, consider the situation shown in Fig. 6 as a typical example. In this situation, most of the localization algorithms based on robust quadrilaterals regard this quadrilateral as a potential source of the flip ambiguity and discards this quadrilateral. As a result, failure in localization of some nodes in the network is often observed. On the other hand, as shown in Fig. 6, this situation does not cause any problem to the inverse Dealunay algorithm. This is due to the bigger margin allowed in checking the Delaunay condition than that for robustness of the quadrilateral.

3 IMPLEMENTATION OF ACOUSTIC RANGING ON A SENSOR PLATFORM (MICA2 MOTE)

Relative distance between sensor nodes is one of the key information for localization. Recently, many algorithms have been developed for ranging. RSSI (Received Signal Strength Indicator) based on attenuation of strength of electromagnetic wave is one of the well-known ranging algorithm for localization with relatively low accuracy applied for sparsely arranged sensor network. On the other hand, for applications with a requirement for precise localization of densely arranged sensor network, more accurate ranging method is utterly needed. For this purpose, ranging based on travelling time of electromagnetic and/or acoustic signals, so called time of arrival (TOA) methods are considered to be effective. Especially, research activities on acoustic ranging are active. This is mainly due to the limited resources in the hardware of currently available sensor nodes which is not enough for implementing ranging devices/algorithms based on TOA of electromagnetic signals. Typical sensor node (Mica2) has only 4kB RAM and 12kB ROM with 7.3MHz microcontroller (Amtel ATmega 128L)[1]. This little capacity shows requirement for algorithms with minimum consumption of memory and computational resources. The ranging algorithm discussed in this section is based on TOA of acoustic signal and implemented in Mica2.

3.1 Basic principle of acoustic ranging

The concept of acoustic ranging is based on measurement of the time of flight (TOF) of the acoustic signal between the signal source and the acoustic receiver. The range is estimated from the time measurement, assuming the speed of sound is known. To accurately measure the TOF of the acoustic signal, time synchronization between transmitter and receiver should be kept. However, accurate synchronization between sensor nodes itself is a topic with numerous ongoing research projects[]. Therefore, we do not try to implement synchronization among sensor nodes. Instead, we make use of the difference between the speed of electromagnetic wave and that of acoustic wave. Since the sound propagates much slower in the air (approximately 10^6 times) than RF signals, TOF can be precisely estimated from the time difference of arrival (TDOA) of simultaneously emitted acoustic and radio signals. This can be achieved by having actuator notify the receiver via a radio message at the same time when the acoustic signal is emitted. The difference of the arrival times of the sound and radio signals is a good indicator for the estimation of TOF.

In the following subsections, implementation of this principle of acoustic ranging on a commercially available sensor node (Mica2) is discussed with a major focus on algorithms for reducing the environment noise and for estimation of TOF.

3.2 Noise Reduction

To locate the beginning of the acoustic signal, we need to increase the SNR (signal to noise ratio) of the samples. Since disturbances such as ambient noise are of Gaussian nature, they are independent for each chirp, whereas the useful signal content will be identical. The acoustic signal used in this research consists of a series of chirps, all of the same length, with difference in intervals of silence in between. As shown in Fig. 7, Delays between consecutive chirps are varied to avoid a situation when multiple samples have the same noise pattern at the same offset, which is a common phenomenon caused by acoustic multi-path effects such as echoes from a wall in indoor environment (see Fig. 8). The length of intervals and the actual emission time of each chirp are preset in sensor and actuator. Once the sensor has received the RF signal from actuator, it controls the sampling timing based on the preset timing described above. The chirps are sampled one by one, then added together and processed as a single sampled signal. Adding together the series of samples improves the SNR by 10log(N)dB, where N is the number of chirps used. Specifically, in our implementation, by using 64 consecutive chirps in an acoustic ranging signal, the SNR is improved by 18dB.

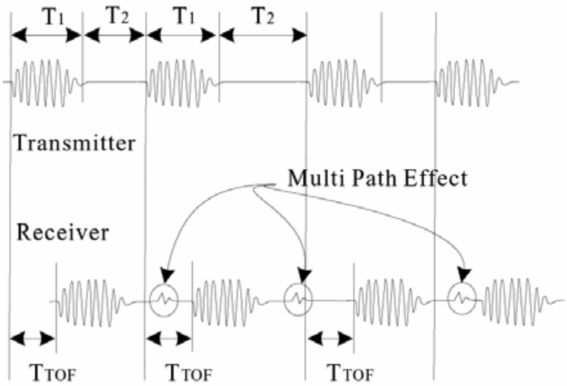


Figure 7 Reduction of the noise by data stacking
T_1=constant, T_2= random

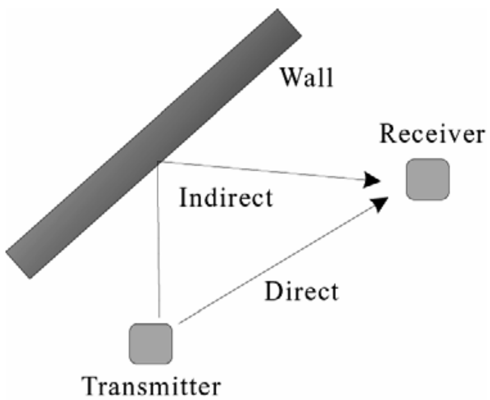


Figure 8 Multi-path caused by reflection

In addition to the external sources of the noise mentioned above (ambient and echo), analogue circuit for microphone and buzzer could contaminate the signal in low frequency range and the internal devices such as crystal oscillator could be the source of high frequency noise. Figure 9 shows the observed digital data after 64 times stacking. Remarkable noise in both low and high frequency range can be observed. To remove these noises, sampled data through ADC (analogue digital converter) is digitally processed by FIR filter and the SNR is improved. For finding integer coefficients of the band pass filter employed in this acoustic ranging application, the real acoustic signal samples containing both chirps and silence obtained by Mica2 is used as input for rectangular window functions. Rectangular window function is used in order to gain the equal weight for 4.0--4.5kHz frequency range of the sounder. By using the rectangular window function, tap number is varied. Figures 10--12 show filtered data with tap number 21, 25 and 35, respectively. We do not observe significant difference among them. Thus, FIR filter with tap number 21 is enough and we implement this filter in Mica2. Knowing the Mica2 hardware limitation, 21-tap FIR filter with integer coefficients in the [-7, 7] interval was chosen.

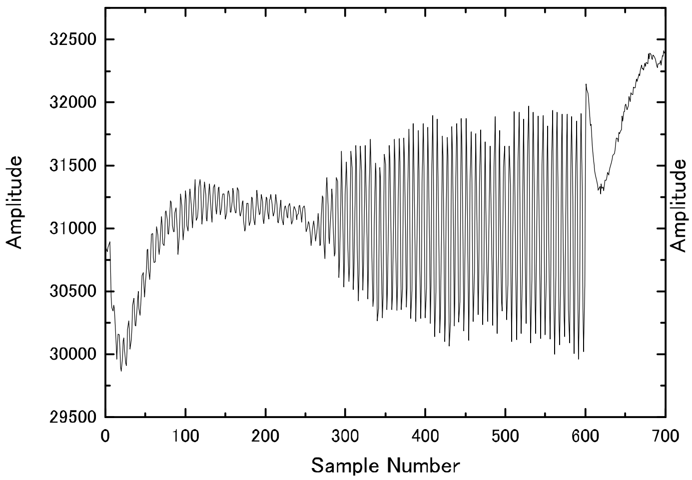


Figure 9 Stacked data without filtering

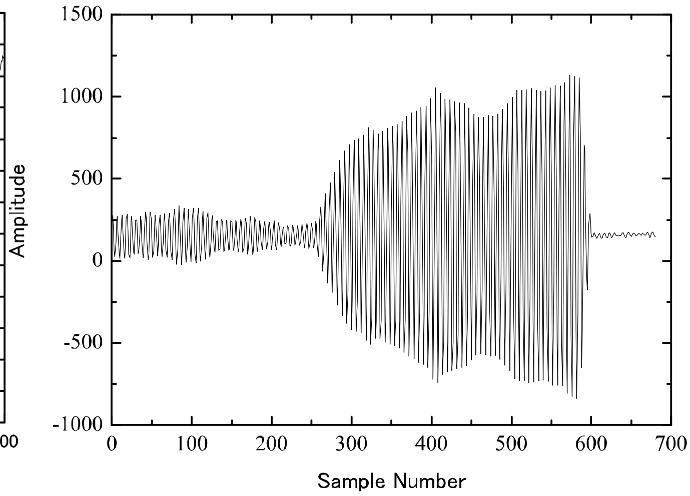


Figure 10 Filtered data with (tap number) =21

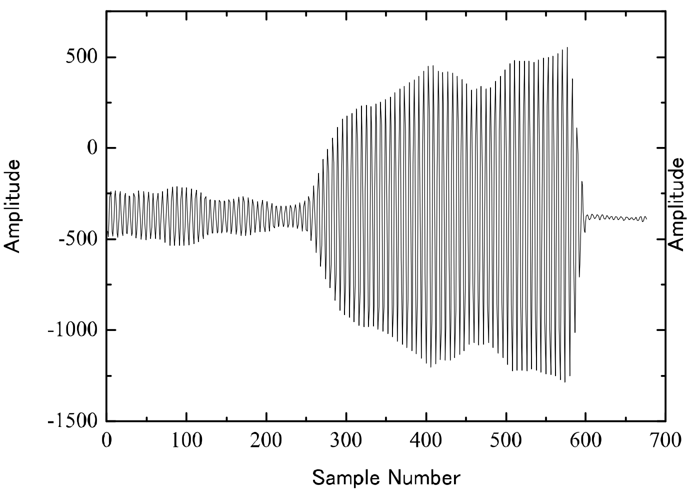


Figure 11 Filtered data with (tap number) =25

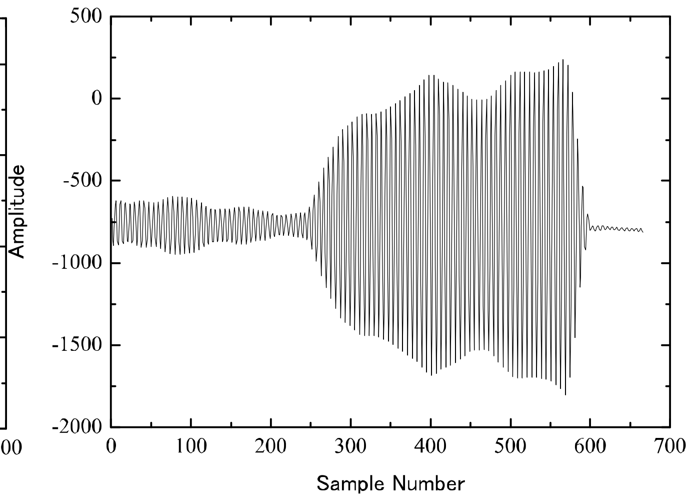


Figure 12 Filtered data with (tap number) =35

3.3 Time of flight estimation

In order to effectively locate the time of arrival, the following algorithm is proposed.

1. Locate all the peaks for the filtered signals.
2. Pick up the maximum amplitude A_{max} from the first 40 samples of the located peaks. For the frequency range between 4.0kHz and 4.5kHz, the 40th sample of the located peaks corresponds to about 100cm distance in the real space. Since the spacing between child nodes are more than 100cm in our application, the maximum amplitude A_{max} is the maximum ambient noise level.
3. Using A_{max} as a reference point, compare the following peaks with it. Locate position Ts where the following peaks are all bigger than A_{max}. This is based on the fact that, then acoustic signal arrives at the sensor, its amplitude is much bigger than the ambient noise after the filtering. Using A_{max} also speeds up the process of detecting the general position of time of arrival.

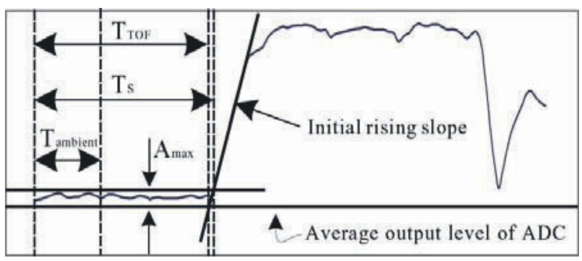


Figure 13 TOF detect algorithm.

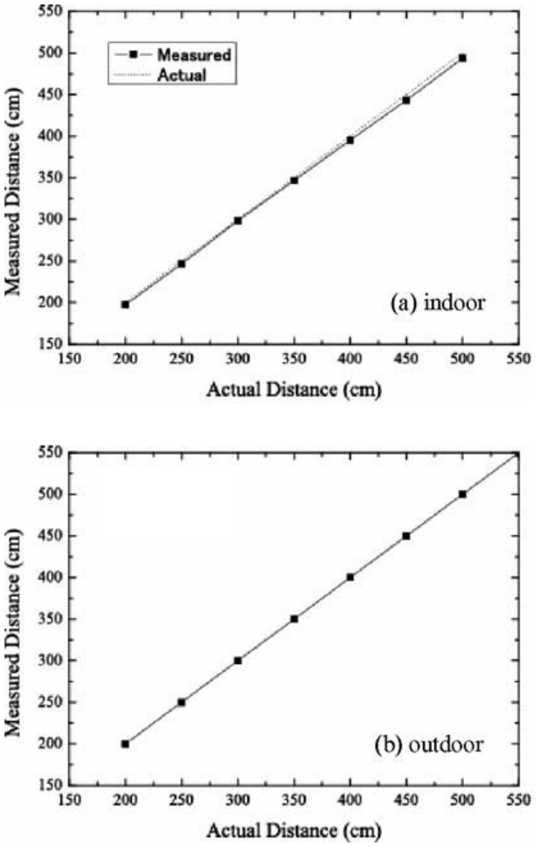
4. Perform adverse look-back to find the exact position of time of flight T_{TOF}. The exact position of time of flight is defined as a point in which the initial rising slope intersects the average output level of ADC, as shown in Fig. 13. Assuming the rising slope linear, a simple Kalman filter is adopted for extrapolating the initial rising slope.

3.4 Experiment for evaluation of acoustic ranging method

The performance of the acoustic ranging method is evaluated in the indoor environment where we observe severe echoes as well as the moderate urban noise environment. Figure 14 (a) and (b) show the comparison between actual and measured distance for the indoor and the outdoor environments, respectively and Fig.15 (a) and (b) show the measurement errors for the indoor and the outdoor environments, respectively. It is clearly visible that the precision of the measurement in the urban noise is better, especially around 300cm. This is because of the fact that we have conducted the indoor experiment in 6m× 6m× 8m room surrounded by hard walls which reflect acoustic wave so that the effect of the echo is severe. However, in any cases, it can be concluded that the accuracy of the present acoustic localization method is with 6cm in terms of standard deviation even in the indoor environment.

4 FIELD EXPERIMENT

Field experiment was conducted to evaluate the performance of the system for localization of sensor nodes based on the inverse Delaunay algorithm and acoustic ranging. Total of 21 sensor nodes (Mica2) with acoustic ranging devices/algorithm discussed in section 3 were spread on the plane (i.e., 2-dimensional arrangement) in a pseudo-random manner as shown in Fig. 16. Each sensor node measured the relative distance among neighboring 4 to 13 nodes depending on the local arrangement of the nodes (i.e., the nodes placed near the boundary measured relative distance among less number of nodes and the nodes at the vicinity of the center of the arrangement measured many). Localization was performed by using measured relative distance among sensor nodes as input data set for the inverse Delaunay algorithm.



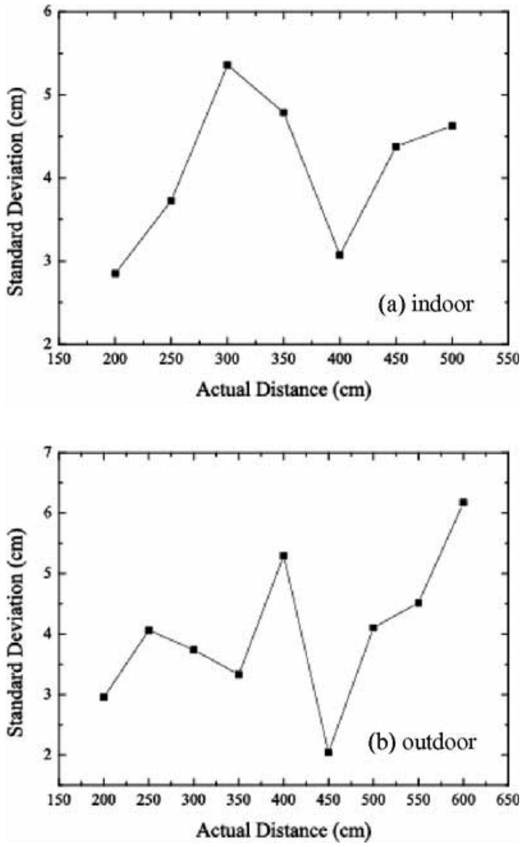


Figure 14 Comparison between actual and measured distance **Figure 15 Measurement error as a function of distance**

Figure 17 shows the estimated and the actual location of the sensor nodes. Since the inverse Delaunay algorithm determines the position of each sensor node upto rigid body motion (translation, rotation and flip) ambiguity as a whole, the global coordinate of the whole network of the sensor nodes is determined by GPS receivers deployed on three anchoring nodes as shown in Fig. 16. The measurement error in the acoustic ranging in this experiment was 3.8cm in average (maximum of 16.8cm) against average distance between nodes of 321.6cm. In spite of this noisy measurement of the relative distance, no node was missed in the localization and the error in the estimation was suppressed in the reasonable range. The error in the estimation of the location of the boundary nodes is relatively large. However, it should be noted that no attempt to redistribute the error throughout the domain, such as Laplacian smoothing or spring relaxation, were not applied in this estimation. No sign of the flip ambiguity, accumulation of the error can be observed. Although more detailed validation such as numerical simulation on the arrangement of thousands of nodes with noise in the ranging as a controlling parameter is required, this experimental result implies the robustness of the system.

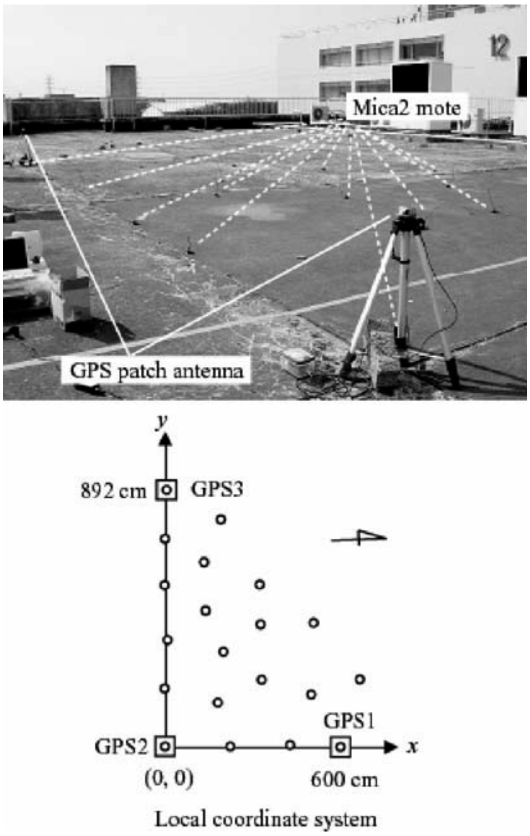


Figure 16 Schematic view of the deployment of sensor nodes

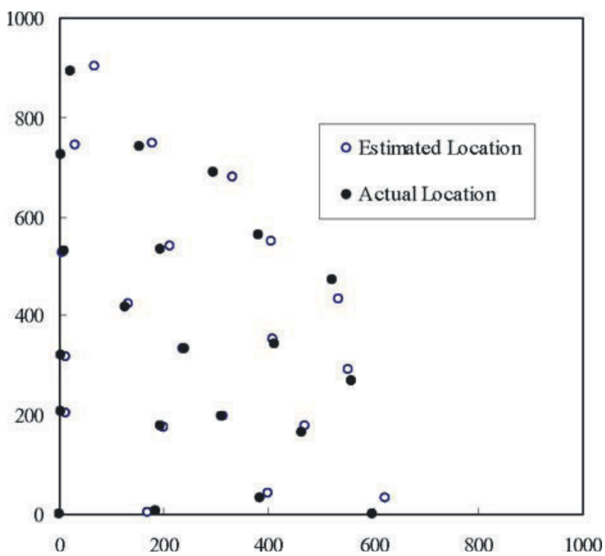


Figure 17 Comparison of the global positions of the nodes localized by our system with the actual positions of the nodes

5 CONCLUDING REMARKS

The inverse Delaunay algorithm and acoustic ranging method for localization (determination of the position of each sensor node) of a large scale sensor network are presented in this paper. Localization is an important issue for applying network sensing technique to the field of civil engineering and or earthquake engineering (e.g., more advanced health monitoring of the infrastructures).

The inverse Delaunay algorithm localizes all the nodes simultaneously, thus, the accumulation of the error in the localization is suppressed. Also, numerical simulation shows the scalability of this algorithm. In all the operation for localization, computational complexity is kept O(n). Noise tolerant acoustic ranging algorithm that employs digital signal processing techniques is implemented in an off-the-shelf sensor platform (Mica2). Experiments show that this acoustic ranging algorithm is sufficient to give average range estimation error below 10 cm. Together with this acoustic ranging algorithm, the inverse Delaunay algorithm is also implemented in Mica2. This enables automatic localization of sensor network. Field experiment was conducted to evaluate the accuracy of the localization of the system. Results from this experiment validate the robustness of the proposed method.

As future work, i) experimental validation of the robustness of the method against errors using hundreds of sensor nodes, ii) extension of the algorithm to 3D deployment of the sensor nodes, iii) improvement of the accuracy of the acoustic ranging and iv) measurement of the actual infrastructures, could be listed.

6 REFERENCES

[1] Hill J & Culler D. (2002) Mica: A Wireless Platform for Deeply Embedded Networks. *IEEE Micro*, 22(6), pp.12-24.

[2] Hofmann-Wellenhof B, Lichtenegger H & Collins J. (2001) *GPS, Theory and Practice*. Springer Wien NewYork.

[3] Moore D, Leonard J, Rus D & Teller S. (2004) Robust Distributed Network Localization with Noisy Range Measurements. *Proceedings of the Second ACM SenSys*.

[4] de Berg M, van Kreveld M, Overmars M & Schwarzkopf O. (1997) *Computational Geometry, Algorithms and Applications*. Springer-Verlag Berlin Heidelberg.

A RISK MANAGEMENT TOOLKIT FOR INTEGRATED ENGINEERING ASSET MAINTENANCE

V Lakshmi Narasimhan

School of Electrical Engineering and Computer Science The University of Newcastle, NSW 2308, AustraliaEmail: lakshmi.narasimhan@newcastle.edu.au, URL: www.cs.newcastle.edu.au/~narasimhan

"A ship in harbour is safe - but that is not what ships are for"

~John A. Shedd, *Salt from My Attic*

Abstract: This paper presents the design of a risk management toolkit that facilitates maintenance of integrated engineering assets over a distributed operational environment. Key aspects of the tool includes impact profiling (i.e., the impact of loss of a particular system or sub-system), parametric modelling of the system-as-a-whole and business case generation. A variance based model has been developed for risk assessment. The system provides number of other features including, but not limited to, report generation, alert and flash messaging, testing/testability considerations and other such records. The system also offers a limited degree of visualisation capabilities in order to view risks at various layers and types.

Key Words: Risk management, Impact profiling, Parametric modelling, Risk visualisation, Variance modelling.

1 INTRODUCTION

Engineering systems typically last over 40 years, be they chemical plants, manufacturing pipelines or process industry systems. While the design of such systems are challenging, design for maintenance is something rarely attempted, unless the systems are mission-critical (e.g., heart pace makers, satellites, etc). However, maintenance takes up more than 75% of the costs of the overall system life cycle costs in almost all engineering projects [10, 33]. Indeed this is the case even with software systems too!

Even though technology changes with time and changes in design and development processes, the downstream issue of overall system/plant maintenance is fraught with considerable risks of various types. Further, the nature and severity of these risks usually compound during the maintenance phase and the costs to address or ameliorate or mitigate these risks also considerably increase – in some cases exponentially during the maintenance phase. The problem is very much akin to noting and trying to fix a stich in a sweater after it has been completely knitted [2], except by then one may have to un-do many stages of stiches in order to correct the original error. The cost of fixing an error can be 100 times as high as it would have been during the development stage and the associated risks in testing and verifying the correct operation of the modified system is also order of magnitude higher than during the original development of the system.

In this paper, we identify various types of risks during the maintenance phase through parametric attributes. For this purpose, we note the definition of the term risk as, "*Risk is the net negative impact of the exercise of a vulnerability, considering both the probability and the impact of occurrence [11]*". They include risks in testing (during maintenance), on-going risks to business and modelling the risk management capability. We develop a variance model for risk management, mitigation and minimisation during the maintenance phase and we believe that this model serves as a step in the right direction towards addressing risk related issues in large organisations.

The rest of the paper is organised as follows: the next section presents more specific details on the risks in integrated systems development, while section 3 deals with factors in engineering risk assessment and related models. Section 4 is on the development of a variance model for evaluating systems risk assessment, followed by a discussion on this model in section 5. The conclusion summarises the paper and provides further research directions.

2 RISKS IN ENGINEERING SYSTEMS DEVELOPMENT & MAINTENANCE

Typical System Development Life Cycle (SDLC) [30, 32] has a number of steps (i.e., system analysis, design, pilot development, implementation, testing and maintenance) and the steps therein vary to some degree depending on the type of system or plant being developed (e.g., manufacturing systems, chemical process engineering systems, etc). Naturally, several types of risks occur during each stage of the SDLC.

Unfortunately, the impact of costs and risks involved in plant maintenance has not permeated to financial managers, who still perceive maintenance as a necessary evil, as opposed to affirmative good. Indeed the Maintenance Hall of Shame [2] showcases several projects, which have been totally abandoned after spending budgets ranging from $11 million to $4 billion. These projects in the OECD world include BHP Billiton's Iron Brickett facility, FBI's Virtual Case File [1, 4] and Anaconda Nickel in Western Australia.

Maintenance occurs in four ways [8, 9]: corrective maintenance, adaptive maintenance, perfective maintenance and preventative maintenance. The first two issues concern fixing of systems, the latter two issues relate to enhancing the system. The corrective approach is intended to defect identification and removal of defects, the adaptive approach manages changes resulting from system changes, the perfective approach handles changes resulting from user requests and the preventative approach deals with changes made to the plant to make it more maintainable. Risks arise at all of the maintenance phases and two impact factors that we are interested in this paper include risks to business and risks to re-testability.

A short and quick literature survey relevant to this paper is as follows: Yassine et. al. [29] provide an information structured approach to designing systems, Gits [20] provides a comprehensive literature review of maintenance concept for various systems, Stewart, et. al. [22] detail probabilistic risk assessment methods and Dunthie, et. al. [23] provide risk-based approaches to maintenance. Edwards [21] provides a specific case study in the construction industry. Risks related to software systems, related frameworks [16], testability [24], cost-benefits [19], choice of test tools [26] and typical mistakes [17, 18] have also been extensively studied. The latter set of references provide an interesting dimension to risk in the all ubiquitous computer software-based systems, which will soon encompass almost all systems. Various IEEE and NIST Standards [8, 11] deal with Software Maintenance issues and analysing problems during maintenance and provides a way to go about addressing these issues.

Charette [2] list several major factors for the failure of such projects. We have amended the original list in order to generate the list of risks that occur during the plant maintenance phase (Table 1).

- Unrealistic or unarticulated project goals
- Inaccurate estimates of the needed resources for maintenance
- Badly defined maintenance requests
- Poor reporting of the system status under maintenance
- Unmanaged risks during maintenance
- Poor communication among users and maintainers
- Use of immature technology in implementation
- Complexity of the system
- Poor maintenance practices, processes and procedures
- Stakeholder politics, and
- Commercial pressure
- Lack of or improper deployment of maintenance Standards

Table 1: List of Typical Risks During Maintenance Phase

The impact of the process chosen and the cultural disposition of the technical people[1] is profoundly felt during maintenance. For instance, Cusumano et al report [5-7] that Japanese companies have applied extensive quality control procedures that they employ for manufacturing to several other industries, including software development also. As a consequence, the reported median number of defects in Japanese IT projects is one fourth of the corresponding numbers in US IT projects. When measured over 100 projects, Cumano et al [7] report that Japanese adherence to rigid software processes produced 0.02 defects per 1000 source lines of code, compared to 0.40 for corresponding US projects. Further, their rigid processes involving considerable amount of comprehensive documentation allows considerable degree of code re-use also. All these issues have a profound impact on maintenance of software systems.

3 FACTORS IN PLANT RISK ASSESSMENT DURING MAINTENANCE

Issues that occur during maintenance include the following:

- Changing priorities
- Testing methods
- Performance measurement
- Incomplete or non-existent system documentation
- Adapting to changing business requirements
- Backlog size
- Measurement of contributions
- Low morale due to lack of recognition or respect
- Lack of personnel, especially experienced
- Lack of maintenance methodology, standards, procedures and tools . . . . .

Obviously, the number and types of unresolved issues determine the level of risk [12, 25] posed by the project. However, one can use high-level estimation outputs to make critical decisions on risks only if the underlying processes can be trusted. When parametric measurements of risk prove inadequate, risks can be simply stated in terms of issues that remain to be resolved yet and their perceived impact to the plant. For example, a good indication on the maintainability of a plant can be inferred by the work of Oman [14], whose key points are brought up in Table 1.

Table 2: Effects on Maintainability of Asset Properties

Our risk models are adaptations of the Function Point models employed in software cost and effort estimations and the works of Oman [14]. The overall system interaction model (adopted and modified from the MIT-90 model for assessing and achieving change [27] and also several available at [13]) for software maintenance is captured in Fig.1, wherein we identify the following elements as having profound impact on the overall impact contributing to the vulnerability of an organisation's Asset Maintenance system:

---

[1] Glazewski [31] makes the following observation about risk management: *We have all heard the saying, "Give a man a fish, and you feed him for a day. Teach a man to fish, and you feed him for a lifetime." Let me revise that from a risk management standpoint: "Put out a manager's fires, and you help him for a day. Teach a manager fire prevention, and you help him for a career." If a manager understands good risk management, he can worry about things other than fire fighting.*

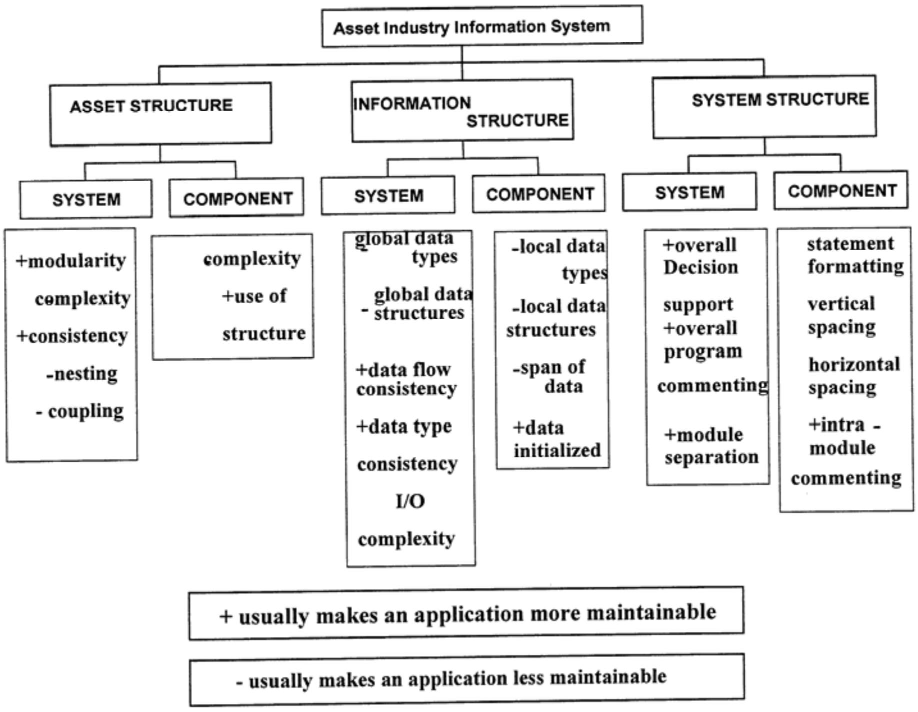


Table 2: Effects on Maintainability of Asset Properties

- Technology, which changes rapidly
- Maturity of skill-base, which is vital for maintenance
- Maturity of plant sub-systems and related entities, which is critical to the stability (which in turn depends on the percentage change in the system per day) and,
- Threat or risk to/from individuals to the IT system. Note that this is not a malicious threats, but more along the lines that "what if a key personnel leaves the organisation, then what will be the impact on plant maintenance"

One can notice that different combinations of factors influence: i) the organisational culture, ii) external technical environment and iii) external socio-economic environment. People involved in the maintenance of plants have to deal with these issues pragmatically and on a continual basis (more on our model for maintenance will be described in a forthcoming paper).

We also define the following factors as contributors to risk during software maintenance and these factors are elaborated in Tables 3-6:

- α-Risk Factors, which are system related
- β-Risk Factors, which are process related
- γ-Risk Factors, which are practice related
- δ-Risk Factors, which are testing related

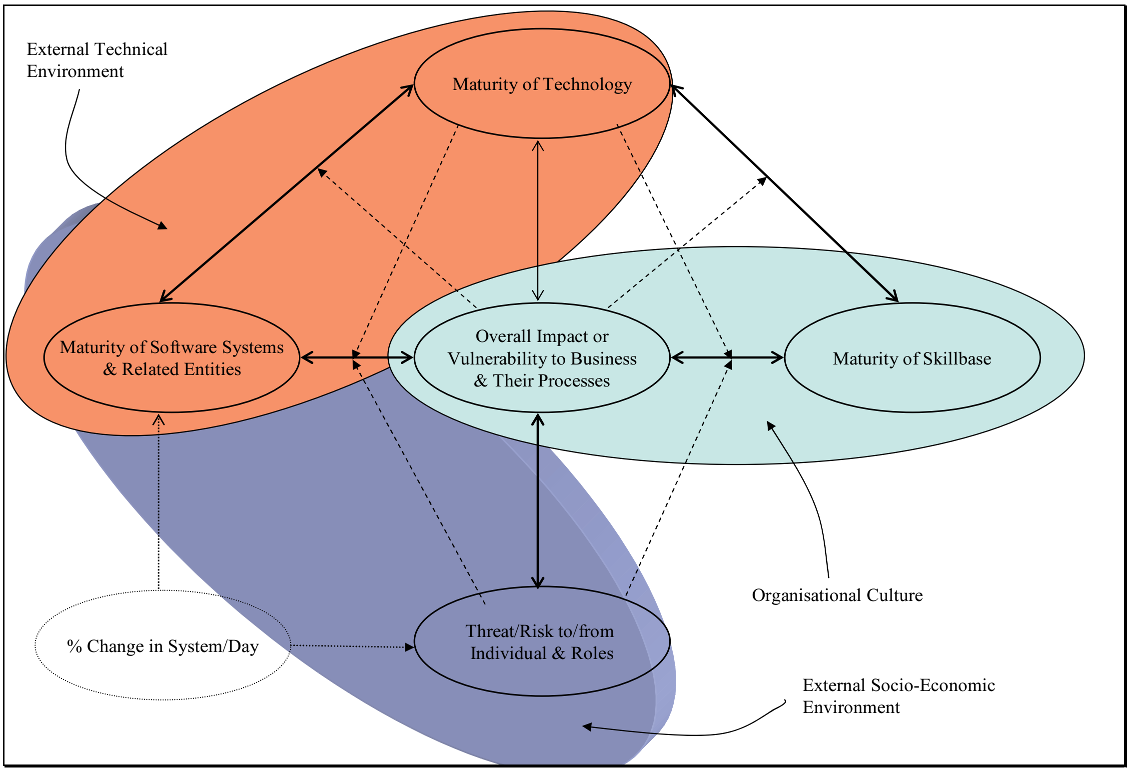


Figure 1: Overall system interaction model for Asset Maintenance

α - Risk Factors	Measurement Mechanism
Size of system	for every 10 sub-systems, add 1 point up to 8 points; for every 10 sub-systems thereafter, add 8 points
Number of systems affected	Actual value
Mean time to test	Actual value
Mean time to fix a bug	Actual value
Average cost per unit test	Actual value

Table 3: α - Risk Factors (Code Related)

β - Risk Factors	Measurement Mechanism
Volatility of development process	no process (=10), *ad hoc*, well-defined, repeatable (=1)
Perceived impact to/on organisation	On a non-linear ***multiple*** of 1 (low), 3, 7, 10, 20 (high)
Degree of clarity of plant requirements	On a scale of 1 (good clarity) – 10 (poor clarity)
Degree of clarity of testing requirements	On a scale of 1 (good clarity) – 10 (poor clarity)
Degree of reusability requirement	On a scale of 1 (low need) – 10 (high need)
Degree of availability of reusable test rigs	On a scale of 1 (poor availability) – 10 (good availability)
Degree of maintainability of reusable test rigs	On a scale of 1 (good maintainability) – 10 (poor maintainability)
Degree of comprehensiveness of the test tool & environment	life cycle toolkit (=1), … , *ad hoc toolkit* (=10)

Table 4: β - Risk Factors (Process Related)

γ - Risk Factors	Measurement Mechanism
Adherence to Standards (*a la* training time)	On a non-linear ***multiple*** of 1 (low), 3, 7, 10, 20 (high)
Time to test	Average time to test a sub-system – actual value
Remaining time to deploy	Actual value
(Efficacy) weighted number of testers	on a non-linear ***multiple*** of 1 (low), 3, 7, 10, 20 (high)
Degree of (real time) performance requirement	On a scale of 1 (low) to 10 (high)
Clarity of performance requirement	On a scale of 1 (low) to 10 (high)
Degree of data/access security constraints imposed	On a scale of 1 (low) to 10 (high)
Degree of code-level migration requirement	On a scale of 1 (low) to 10 (high)
Degree of plant-level migration requirement	On a scale of 1 (low) to 10 (high)
Percentage of full load testability requirement	Actual value

Table 5: γ - Risk Factors (Practice Related)

3.1 The δ-Risk Factors

- Number of users that use the plant
- Number of control sub-systems in the plant
- Number of external sub-systems accessible through the plant
- Number of inspections conducted per day
- Total available plant performance
- Number of approved changes per day
- Number of maintenance people
- Number of maintenance centres in the organisation (there could be a "cultural difference" between the testing centres!)
- Number of total sub-task in testing/maintenance

Table 6: δ - Risk Factors (Testing Related)

4 VARIANCE MODEL FOR ASSET MAINTENANCE RISK ASSESSMENT

We define several risk factor metrics as noted in Table 7, before we develop the variance model.

Risk Factor Metrics	Definitions
Coverage	Percentage amount of the plant tested after a maintenance cycle
Impact	Degree of effect of the maintenance operation from one cycle to another, which can be positive or negative
Time2Fix	The average time taken to fix a reported bug (does not include bug detection time)
Exposure	Degree of revelation of faults after a particular maintenance cycle that gets exposed to outside of the organisation

Fault Likelihood	Probability of a fault occurring
Consequence	The consequence of failure of a particular sub-system
Degree of Adherence	Degree of adherence to procedures, Standards or process and their perceived veracity and effectiveness towards the entire maintenance process
Probability of loss of Key People	The impact of the loss of key people
Remaining Time	The remaining time before the proposed re-start of the plant

Table 7: Risk Factor Metrics and Their Definitions

We now define the following metrics as noted in Table 8. Note that each one of the functions (F#(t)) below is unique and only the key parameters from each of the risk factors from Tables 3-6 are included in the actual calculation of the metrics; the remaining parameters are ignored. The selection of the right type of parameters is based on experience, application requirement and context in which the measurement is sought. Further the function F#(t) is a measure of the variance or changes in the parameters as the entire plant goes through various stages of its maintenance cycle at any given time t.

Risk Factor Metrics	Nature of Measurement
Coverage	F# (α-Risk Factors)
Impact	F# (δ-Risk Factors)
Time2Fix	F# (γ- and δ-Risk Factors)
Exposure	F# (α- and γ-Risk Factors)
Fault Likelihood	F# (α , γ and δ-Risk Factors)
Consequence	F# (δ-Risk Factors)
Degree of Adherence	F# (β-Risk Factors)
Probability of loss of Key People	F# (β and δ -Risk Factors)
Remaining Time	F# (γ-Risk Factors)

Table 8: Risk Factor Metrics and Their Measurement

The metrics given in Table 8 can be variably evaluated depending on which of the listed risk factors are considered to dominate a particular plant. As a further extension, one can define a number of metrics for the following terms also:

- <u>Impact Metrics:</u> Integrity, Availability, Confidentiality, Vulnerability, Threat-source Identification, Threat-action Identification.

- <u>Mitigation Action & Strategy Metrics:</u> Risk Awareness, Risk Assumptions, Risk Avoidance, Risk Limits, Risk Plans, Risk Acknowledgement, Risk Transfer or Risk Outsourcing.

- <u>Cost-benefit Metrics:</u> for the following actions - Analysis, Assigning responsibility, Priority/Rating generation, Cost of detection, Cost for correction, Targeting, Process/Standard generation, Education & adherence to process/Standards (Training), Standards/process improvement (capability maturity), identifying & evaluating residual risks.

- <u>Incident Management Metrics:</u> Incident reporting, Categorization, Incident Prioritization/Rating, Responsibility assignment, Risk isolation, Impact assessment.

5 DISCUSSION OF THE VARIANCE MODEL

We define several types of metrics for software maintenance, which include the following: risk to testability, risk to business activities and risk to business perception. These are pictorially described in Fig.2 below. The empirical implications of various TAR values are provided in Table 9.

Risk to Testability (R2T) = **A** = Area of triangle {Coverage, Impact, Time2Fix}

Business Perception Risk (BPR) = **B** = Area of triangle {Exposure, Fault Likelihood, Consequence}

Business Vulnerability Risk (BVR) = **C** = Area of triangle {Degree of Adherence, Prob. loss of key people, Remaining Time}

Total Acceptable Risk (TAR) = **A * B * C**

Other factors such as Risk to Business can be in terms of performance, time-to-market (or market conditions) and technology alternatives. For example,

Risk2Business (t) = F3 (c1, c2, c3, c4) / F4 (d1, d2, d3, d4)

C1 = perceived impact on business (visibility, loss of life,)

C2 = degree of Leadership support

C3 = degree of organisational culture to support the values of testing and testability

And, the rest are explained in the forthcoming paper.

Similarly, risks can be quantified or qualified in terms of humans and systems as noted below:

Human Terms (t) = F (easy of training test tool/environment, % change in policies, nature of project, recoverability from errors, # maintenance centres,)

System Terms (t) = F (# of sub-systems, # of controllers, # of transactions/sec, # of refreshes/day, # changes/day, degree of concurrency required,)

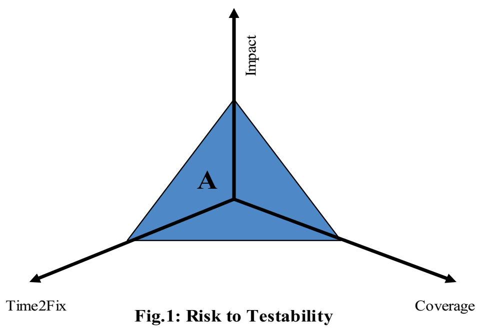


Fig.1: Risk to Testability

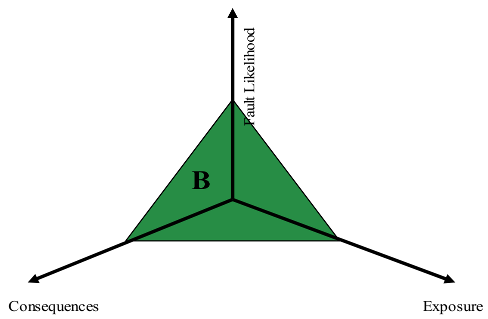


Fig.2: Business Perception Risk

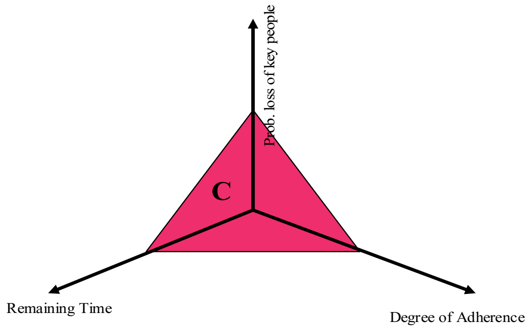


Fig.2: Business Vulnerability Risk

TAR Values	Implications	Recommended Remedial Actions
10	Fatal failures	Immediate attention with highest priority
7	Serious failures	High priority; expect consequences
4	Manageable failures	Good priority; consequences may be serious.
1	Acceptable failures	Low priority, consequences may not be serious
0.1	Virtually impossible	Well tested, rare errors with minor consequences

Table 9: TAR Values and their Implications

5.1 Development of a Risk Measurement Toolkit

We have developed a version of our Risk Measurement Toolkit using Excel Spreadsheet, whose parameters are provided in Table 10, which itself is a modified version of the NIST Standard on Maintenance Risk Analysis Matrix [11]. We also generate some recommendations automatically and further, repositories for the creation and maintenance of a repository for each of the following items:

- Risk watch-list
- Profile of typical errors, bugs, pitfalls and mistakes
- Code patterns for error corrections
- Potential impact profiles

- Key personnel impact
- Knowledge gained from correcting bugs, pitfalls and mistakes

Nature of Risk	Priority	Recommended Control Actions	Planned Controls	Required Resources	Responsible Team or Persons	Start/ End Dates	Estimated Total Costs	Residual Risk	
•									
•									
•									

Table 10: Maintenance Risk Analysis Matrix

5.2 Maturing of the Risk Model - WArm-blooded Risk Management (WARM) Model for Risk Management Maturity

We present a new model, called the WArm-blooded Risk Management (WARM) model (see Table 11), which indicates the degree of maturity of organisations in managing their risk. The WARM model is inspired by the works of Pooch [15], Capability Maturity Model Integration of Humphry [28] and discussions with various industrial experts. At the basic step, all organisations are Cows, where risk management happens by accident of serendipity. As the organisations improve their risk management profiles, they move in their maturity levels from the Cow-level, to that of elephants, hyenas, foxes and then lions, one-level at time. Along with their maturity level upward movement, their capabilities also increase profoundly. At the highest level, organisations will be able to consider their entire risk management life cycle and cost their operations in terms of costs, resources, personnel and performance requirements.

Level	Name	Attributes	Characteristics
1	Cow	Scavengers, nature fed, unfocussed; eat whatever available dead or alive, so long as you can eat!	Risk management happens by accident or serendipity.
2	Elephant	Gulp all, unfocussed irrespective of quality, no particular motivation except to fill-in.	Limited risk management is performed, but just to fill the scripts, with no planning or real load based evaluation.
3	Hyena	Acting as a team on easy target, predatory, fast.	Automated tools are used to perform risk management, but inconsistent processes employed with limited estimate on costs.
4	Fox	Group action, movement, temperate.	Near-production risk management is performed with clearly repeatable processes, control over performance requirements and costs, maintainable risk profiles.
5	Lion	Group, motivation, large-scale movement across entire available space.	Entire risk management life cycle is considered, well-articulated risk profiles maintained, costs, resources and performance requirements controlled.

Table 11: The WARM Model for Risk Management Maturity

6 CONCLUSIONS

The measurement of risks and process that must be placed to eliminate or mitigate the risks, involves the collection of data and decision support systems. In this paper, we have identified various types of risks that occur during plant maintenance. We have parameterised these attributes and developed a variance-based model for their measurement and control. We have also identified several metrics for monitoring risk during maintenance, described a toolkit and discussed the WARM Risk Maturity Model. We are currently working on the design and development of an intelligent agent based framework for comprehensive risk assessment during single and multi-plant maintenance.

7 REFERENCES

[1] H. Goldstein, “Who killed the Virtual Case file?”, IEEE Spectrum, vol.42, no.9, pp.18-29, 2005.

[2] R. N. Charette, “Why software fails” , IEEE Spectrum, vol.42, no.9, pp.36-43, 2005.

[3] P.G. Armour, “To pan, two plans: A planning approach to managing risk”, Comm. of ACM, vol.48, no.9, Sept.2005, pp.15-19.

[4] “Inside Risks”, regular column, Comm. of ACM.

[5] M.A. Cusumano, “The puzzle of Japanese software”, Comm. of ACM, vol.48, no.9, Sept.2005, pp.25-27.

[6] M.A. Cusumano, “Japan’s software factories”, Oxford University Press, NY, 1991.

[7] M.A. Cusumano, A. MacCormack, C. Kemerer and B. Crandall, “Software development worldwise: The state of the practice”, IEEE Software, vol.20, no.6, Nov-Dec. 2003, pp.28-34.

[8] E. J. Braude, “Software Engineering: An Object-Oriented Perspective”, Wiley 2001.

[9] Deklava, “Delphi Study of Software Maintenance Problems”, Proc. of Intl. Conf. on Software Maintenance, 1992, pp.10-17.

[10] “EDS Seminar Notes”, EDS IT Technical Manager, Adelaide. Seminar held at Newcastle in April, 2005.

[11] G. Stoneburner, A.Goguen and A. Feringa, “Risk Management Guide for Information Technology Systems”, NIST Special Publication No.: 800-30, US Department of Commerce.

[12] P. G. Armour, “Project Portfolios: Organizational Management of Risk”, Comms. of the ACM, vol.48, no.3, Mar 2005, pp.17-20.

[13] D. Cho, M.Parlar, "A survey of maintenance models for multi-unit systems", *European J. Operational Research*, Vol. 51 pp.1-23, 1991.

[14] P. Oman and J. Hagemeister, "Metrics for Assessing Software System Maintainability," *Proc. Conf. Software Maintenance*, IEEE CS Press, Los Alamitos, Calif., Order No. 2980-02T, 1992, pp. 337-344.

[15] P. Pooch, “Application Performance – A Risk-based Approach”, Second Test Automation Workshop, Bond University, Gold Coast, Australia, 1-2 Sept., 2005.

[16] C. Nigle, “Test Automation Frameworks”, Available at: http://safsdev.sourceforge.net/FRAMESDataDrivenTestAutomationFrameworks.htm

[17] B. Pettichord, “Seven Steps to Test Automation Success”, Available at: http://www.io.com/~wazmo/papers/seven_steps.html

[18] B. Marick, “Classic Testing Mistakes”, Available at: http://www.testing.com/writings/classic/mistakes.html

[19] C. Kaner, “Quality Cost Analysis: Benefits and Risks”, Available at: http://www.kaner.com/qualcost.htm

[20] C.W. Gits, "On the maintenance concept for a technical system: II. Literature review", *Maintenance Management International*, Vol. 6 pp.181-96, 1986.

[21] L. Edwards, “Practical Risk Management in the Construction Industry”, ISBN: 0727720643, Thomas Telford, 1995.

[22] Mark G Stewart, Robert E Melchers, “Probabilistic Risk Assessment of Engineering Systems”, ISBN: 0412805707, 1997.

[23] Duthie J.C.; Robertson M.I.; Clayton A.M.; Lidbury D.P.G., “Risk-based approaches to ageing and maintenance management”, Nuclear Engineering and Design, Volume 184, Number 1, 1 August 1998, pp. 27-38(12).

[24] K. McMahon, "Risk-Based Testing", Available at: http://www.data-dimensions.com/testersnet/docs/riskbase.htm

[25] Yacov Y. Haimes, "Risk Modeling, Assessment, and Management", ISBN: 0471480487 - Wiley-IEEE, 2004.

[26] M. Fewster and D. Graham, "Choosing a Test Tool", Available at: http://www.grove.co.uk/Tool_Information/Choosing_Tools.html

[27] "Model (MIT90) for Assessing and Achieving Change", http://www.adventengineering.com/why/mit_90/mit_90.htm

[28] W. Humphry, "Capability Maturity Model Integration", http://www.sei.cmu.edu/cmm/.

[29] A. Yassine, D. Falkenburg, K. Chelst, "Engineering design management: an information structure approach", International Journal of Production Research, Volume 37, Number 13 / September 10, 1999, 2957 – 2975.

[30] B. S. Blanchard, "System Engineering Management", John Wiley and Sons, ISBN: 0471291765.

[31] S. R. Glazewski, "Risk Management (Is Not) for Dummies", available at: http://www.bettermanagement.com/library/library.aspx?libraryid=12496

[32] J.N. Martin, "Systems Engineering Guidebook", ISBN: 0849378370 - CRC Press.

[33] D.N.P. Murthy, A. Atrens, J.A. Eccleston, "Strategic maintenance management", Journal of Quality in Maintenance Engineering, Volume 8 Number 4 2002 pp. 287-305.

DEGRADATION MONITORING OF INDUSTRIAL HEAT EXCHANGERS

BR Upadhyaya[a] and E. Eryurek[b]

[a] *Nuclear Engineering Dept., University of Tennessee, Knoxville, TN 37996-2300, USA*

[b] *Formerly with Emerson Process Management, Eden Prairie, Minnesota, USA*

Abstract: This paper presents the results of monitoring device-level and system-level faults in industrial heat exchangers. The device-level faults refer to degradation of sensors and control actuators. In the case of heat exchanger (HX) studies presented here, system-level fault refers to fouling of HX tubing during plant operation. The objective of the research performed at The University of Tennessee was to develop techniques to detect and isolate both sensor faults and fouling-related degradation in a tube-and-shell heat exchanger. This was achieved by monitoring the changes in the relationship among a set of system-bound variables for different fault types. The measurement-based models were established using experimental data of fouling progression in a tube and shell industrial heat exchanger. Direct and inferred measurements of fluid temperatures, flow rates, pressure drops, and thermal resistance were used to perform the fault diagnosis. An approach called the Group Method of Data Handling (GMDH) was developed to characterize the interrelationships among several measurements

Key Words: Heat exchanger fouling, Device faults, Data-based modelling, Diagnostics

1 INTRODUCTION

The issue of on-line monitoring of process units and the ability to detect their incipient failures is gaining high priority in many industrial processes, including power generation systems, chemical industry, pulp and paper industry, and many other process industries. Examples of process units include heat exchangers, feedwater heaters, boilers, coal pulverizers, feed water pumps, electric motors, and others. In addition, the monitoring of field devices such as instrumentation systems, controllers, and actuators is equally important. The combined effort helps in maintenance planning and optimal equipment repair and replacement. The research and the results presented here focused on the development of data-driven empirical characterization techniques with application to industrial heat exchangers. In the case of heat exchanger (HX) studies presented here, system-level fault refers to fouling of HX tubing during plant operation [1]. This is an important concern because of the degradation of thermal performance of these process units. Particulate fouling is considered in this study and is defined *as the accumulation of solid particles suspended in a fluid onto a heat transfer surface. Suspended particles can be ambient pollutants (sand, silt, clay), upstream corrosion products, or products of chemical reactions occurring within the fluid.* This was achieved by monitoring the changes in the relationship among a set of system-bound variables for different fault types in a tube-and-shell heat exchanger. The measurement-based empirical models were established using experimental data of fouling progression in the heat exchanger. Direct and inferred measurements of fluid temperatures, flow rates, pressure drops, and thermal resistance were used to perform the fault diagnosis. An approach called the Group Method of Data Handling (GMDH) [1] was developed and applied to the measurements. The GMDH performs a prediction of the process variables.

The current work is based on several earlier works on heat exchanger fouling [2-5]. We used heat exchangers with multiple tubing, and performed on-line measurements of fluid temperatures and flow rates during the entire fouling progression. In addition to experimental data, steady-state heat transfer models were developed to complement the information set, especially in cases where plant data may not be initially available.

The paper describes the analytical approach for calculating the changes in the thermal resistance, the development of the heat exchanger experiment, and the results of monitoring tube fouling and sensor degradation in the heat exchanger system. The results indicate the feasibility of applying this approach for monitoring industrial heat exchangers.

2 THEORETICAL STUDY OF THE EFFECT OF FOULING AT DIFFERENT TUBE LOCATIONS

2.1 Tube Fouling and Thermal Performance

The foulant deposition on tube surfaces results in the degradation of thermal performance in heat exchangers, fossil boilers, and nuclear plant steam generators. Furthermore, tube fouling results in a decrease in the tube internal diameter and consequently an increase in the pressure drop between the inlet and the outlet. For this reason, heat exchangers and steam generators are designed with excess heat transfer capacity in order to offset the effect of fouling. The heat transfer resistance increases with fouling and the overall heat transfer coefficient decreases. This parameter is critical in planning unit servicing and tube plugging, if necessary. Limits on the percent of tubes plugged must be established so that the process unit would perform without losing considerable thermal performance.

2.2 Fouling Behaviour at Different Tube Locations

An attempt has been made to understand the effect of fouling on the fluid temperatures at different locations along the tube. This simple model does not consider the influence of fluid temperature itself on the fouling rate. The following energy balance equation is used to characterize the steady-state heat transfer.

$$\dot{Q} = UA\Delta T_{LMTD} = \dot{m}_h C_h \Delta T_h = \dot{m}_c C_c \Delta T_c \quad \textbf{(1)}$$

$\dot{Q}$ is the heat transfer rate, A is the surface area, and U is the overall heat transfer coefficient. ΔT_{LMTD} is the logarithmic mean temperature difference (LMTD) for the heat exchanger. Note that the subscripts *h* and *c* indicate hot side (tube side) and cold side (shell side) fluids. The LMTD is defined as

$$\Delta T_{LMTD} = \frac{\Delta t_1 - \Delta t_2}{\ln(\Delta t_1 / \Delta t_2)} \quad (2)$$

For a parallel flow heat exchanger

$$\Delta t_1 = t_{h,in} - t_{c,in} \,;\; \Delta t_2 = t_{h,out} - t_{c,out} \quad (3)$$

For a counter flow heat exchanger

$$\Delta t_1 = t_{h,in} - t_{c,out} \,;\; \Delta t_2 = t_{h,out} - t_{c,in} \quad (4)$$

The model used was for a parallel flow HX to match with the experimental device. The flow rates on both the tube and the shell side were kept constant and a uniform fouling per unit area was assumed. Figure 1 shows the cold-side (shell side) coolant temperature distribution along the HX length. The graph indicates that the final cold-side outlet temperature attains the same value independent of the location of fouling. This is also the behaviour of hot-side (tube side) fluid temperature. Thus, the tracking of only the outlet temperature is not sufficient to determine the location of fouling. In practice, it is necessary to measure the temperature profile along the HX tube length by distributed sensors.

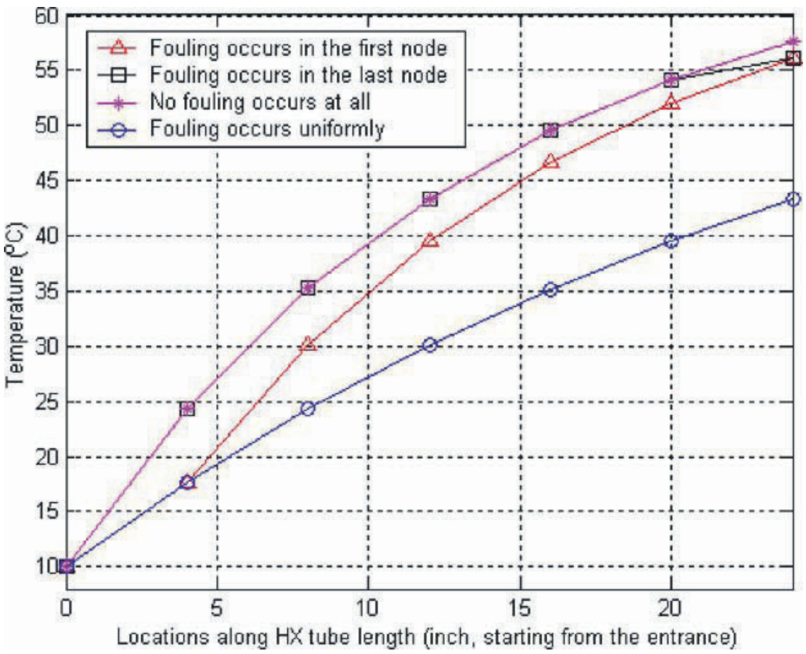


Figure 1. Cold-side coolant temperature distribution along HX tube length.

3 HEAT EXCHANGER SYSTEM DEGRADATION MONITORING

This study has focused on tracking fouling-related degradation and faults in process sensors. The process sensors of interest are measurements of fluid inlet and outlet temperatures, and differential pressure across the tubing. It is important to understand that changes in thermal resistance across the tubing may be caused by fouling, fluid flow rate fluctuations, and/or other factors. In order to separate the process effects from equipment degradation it is necessary to establish a database of normal operating conditions. In the absence of such knowledge, the desired information could be generated using physics models. The data generated from a physics model is then used to develop empirical models for future use.

The following procedure was developed to detect and isolate HX system degradation based on our knowledge of the system physics and the experimental observation.

3.1 Development of the Knowledge Base

Process data are acquired at different operating conditions. If a complete set of data is not available in the beginning, the plant database is expanded by appending information generated from the physics model. As the plant operation progresses, the need for simulated information may be reduced. The normal database is then characterized by building multivariate empirical models (GMDH models described in Section 5). These models are then used to predict one or more process variables using plant measurements.

3.2 Sensor Fault Detection

As the HX fouling progresses, this results in a change in the measured fluid temperature. This is a normal behaviour of the sensor. In this case the temperature profile follows the change in thermal resistance. Assuming that temperature measurements have redundant sensors, a cross check is performed to determine the inconsistency among the sensors. If this is detected, the faulty sensor may be isolated by verifying the heat balance by repeated computation of the heat transfer rate, $\dot{Q}$, using different combinations of hot-side and cold-side temperatures.

$$\dot{Q} = \dot{m}_h C_h \Delta T_h = \dot{m}_c C_c \Delta T_c \tag{5}$$

If sensor redundancy is not available, empirical prediction models may be used to detect one of the four defective temperature sensors: $t_{h,in}, t_{h,out}, t_{c,in}$, and $t_{c,out}$. Multiple prediction models are needed in this case. The predicted values are compared with actual measurements and the resulting residual patterns are evaluated for suspect sensor.

3.3 Heat Exchanger Fouling Detection (HXFD)

When step 2 described in Section 3.2 is successfully completed, HX fouling can be monitored using several parameters: percent change in the thermal resistance, changes in the prediction error of a selected temperature measurement, and differential pressure on the tube-side of the HX. Generally, all these parameters have similar behaviour. Calculation of thermal resistance change is performed by directly using the temperature and flow measurements. The overall thermal resistance of the heat exchanger is calculated as [1]

$$R = \frac{1}{UA} = \frac{\Delta T_{LMTD}}{\dot{m}_h C_h \Delta T_h} = \frac{\Delta T_{LMTD}}{\dot{m}_c C_c \Delta T_c} \tag{6}$$

$\dot{m}_h, \dot{m}_c$: Fluid Flow rates on the hot side and the cold side, respectively

C_h, C_c: Fluid heat capacity on the hot side and the cold side, respectively

$$\Delta T_h = t_{h,in} - t_{h,out}$$

$$\Delta T_c = t_{c,out} - t_{c,in}$$

The thermal resistance generally follows a first order asymptotic model. When fouling is predominant, both the pressure drop and the fluid outlet temperature prediction error follow a similar behaviour. Thus all the three parameters have similar time characteristics. The approach described here (see Ref [7]) is capable of detecting and isolating multiple faults in the heat exchanger system.

4 EXPERIMENTS ON PARTICULATE FOULING OF A TUBE-AND-SHELL HEAT EXCHANGER

4.1 Experimental Loop

The approaches developed for HX fouling degradation monitoring were implemented using a small-scale tube-and-shell heat exchanger. Figure 2 shows a schematic of the HX experimental loop. The process measurements include fluid inlet and outlet temperatures, differential pressure (DP) in the tube, and flow rates on the tube-side and on the shell-side. All the measuring devices were calibrated and the data acquisition was performed using the LabVIEW software.

The demonstration of fouling monitoring by tracking the overall thermal resistance was accomplished by running the experiments for different particulate concentrations and for different tube-side water flow rates. The foulant was a clay-like substance that has the trade name KAOLIN.

4.2 Particulate Concentration and Tube Fouling Degradation

Three experimental runs were made to study the effect of particulate concentration. Each experimental run was about 80-90 hours long. All the data were acquired continuously during the run. At the end of each run, the HX end plates were removed and all the 31 tubes were thoroughly cleaned with brushes and flushed with clean water. The experiment was now run with clean water to verify the normal value of the thermal resistance. If it matches with the nominal value, this is an indication that the change in the thermal resistance is due to tube fouling. The fluid flow rates were kept constant during all the runs. The following three concentrations of KAOLIN were used: 70 gm (1,647 ppm), 90 gm (2,118 ppm), and 110 gm (2,588 ppm). KAOLIN was mixed into the water tank from which the water was circulated through the loop.

Figures 3-5 show the measurements of the process variables, percent change in the thermal resistance during the experimental period, and the pressure drop between the inlet and the outlet headers on the tube side of the HX, respectively, for Run # 3 with 2,588 ppm of KAOLIN on the tube side. Figures 3 and 4 indicate that the time behaviours of the thermal resistance and the pressure drop follow approximately a first order trend, achieving asymptotic values at approximately 120 hours into the test run. It was observed that these values increase with particulate concentration, and so does the time to achieve asymptotic value. At the end of the experiment, the HX tubes were cleaned and the experiment was run with pure water to verify the initial value of the thermal resistance. Figure 6 shows a photograph of the HX tubing with the cleaning brush at the end of the fouling experiment.

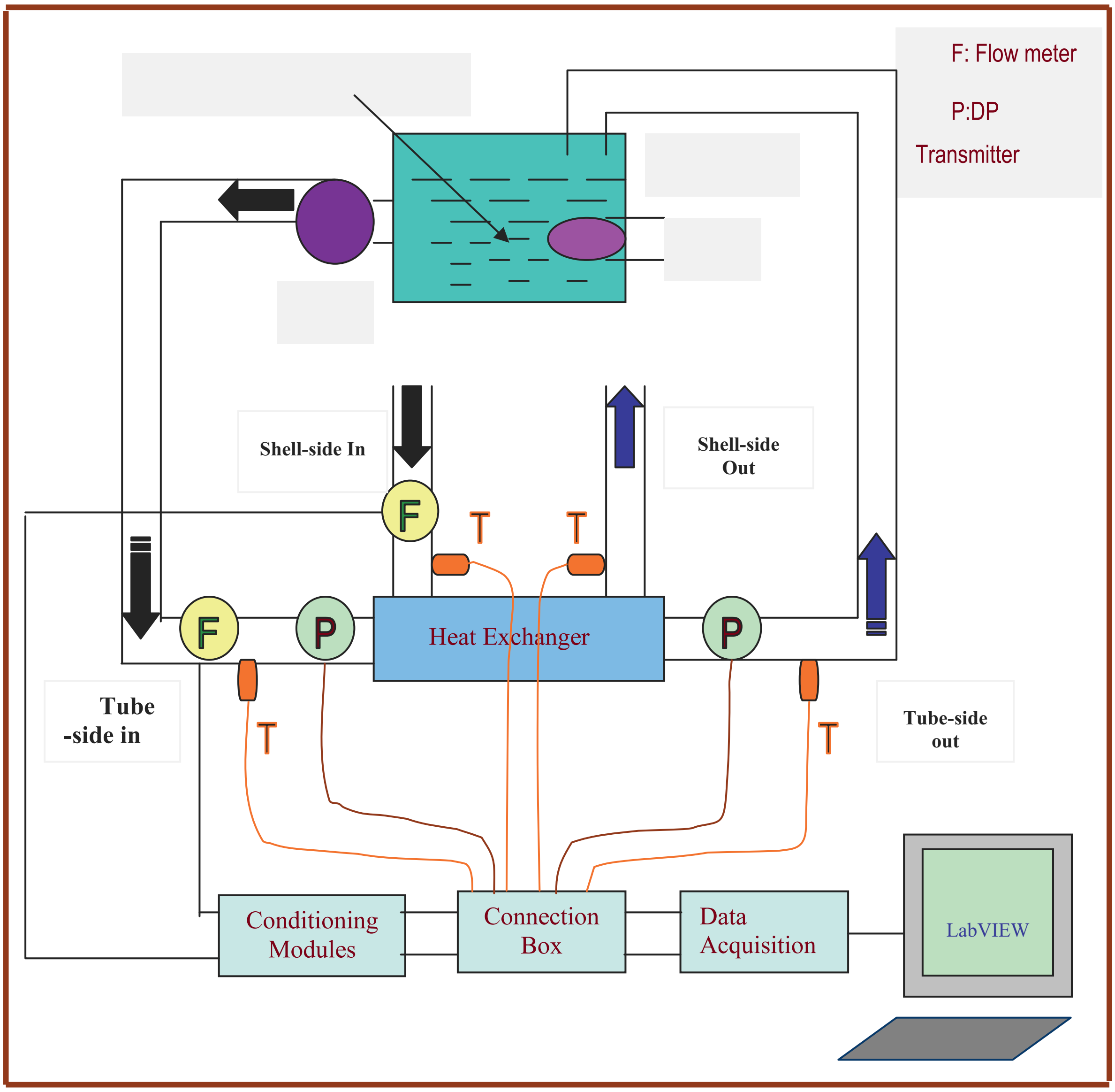


Figure 2. Experimental loop used for fouling studies. All the measurement points are shown.

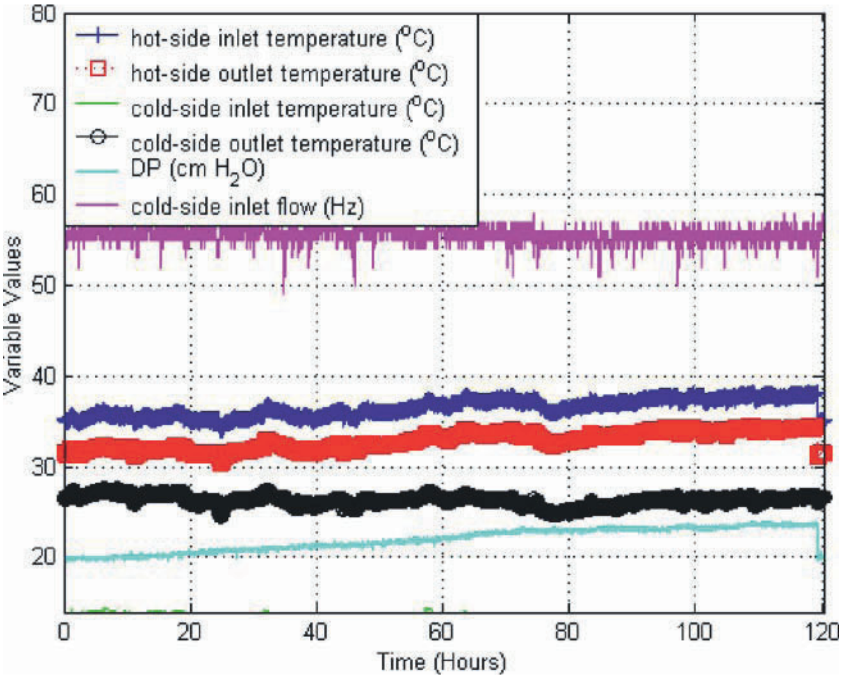


Figure 3. Measurements of cold-side fluid flow rate, temperatures, and differential pressure for Run # 3.

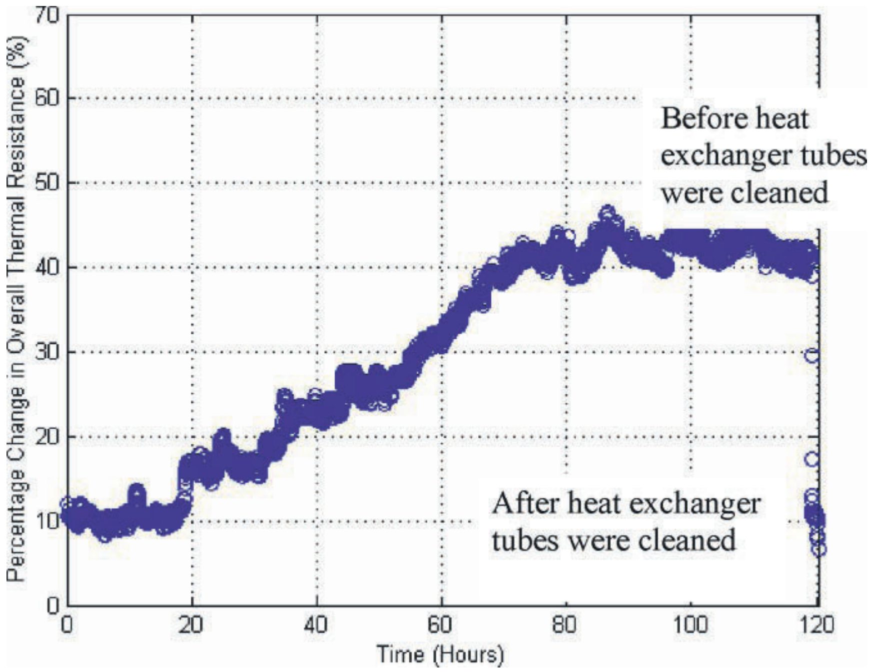


Figure 4. Trending of the calculated overall heat transfer resistance for Run # 3.

5 GROUP METHOD OF DATA HANDLING (GMDH) TECHNIQUE FOR PROCESS CHARACTERIZATION

5.1 Summary of the GMDH Approach

In the absence of a detailed physics model to monitor the cause and effect relationship among system state variables, a multivariate empirical model may be developed using operational data. Even though these models do not take into account all the state variables defined by the physics, they are sufficient to track the changes in the process itself and to monitor the degradation of sensors and field devices. Such an approach is highly efficient and has seen many applications [8, 9]. These empirical models range from simple linear, single-variable to nonlinear, multivariate forms. A proven technique for a general characterization of multiple measurements is the Group Method of Data Handling (GMDH) [6, 10].

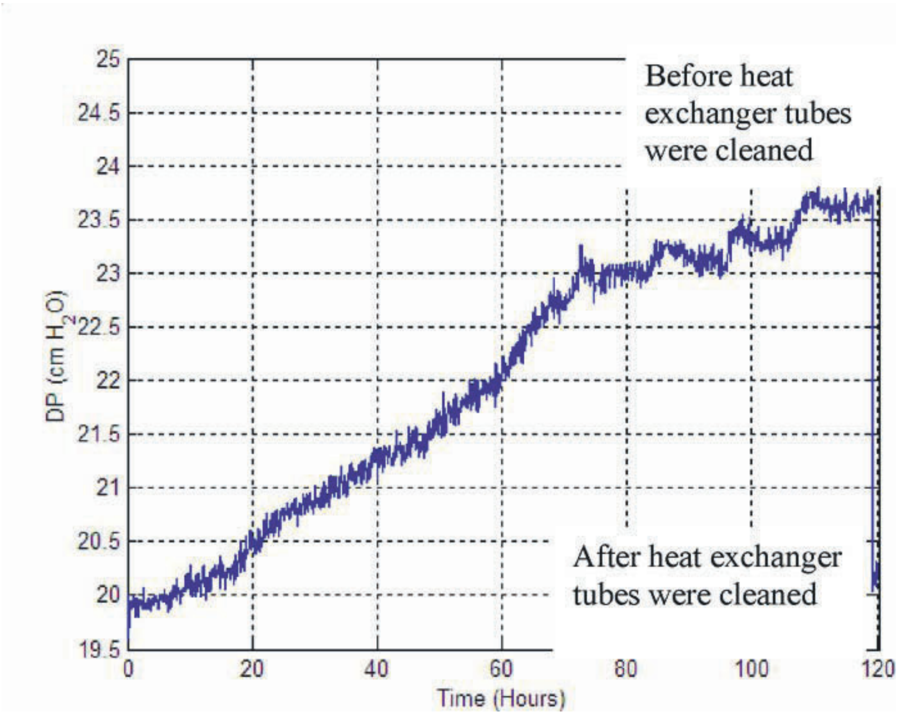


Figure 5. Trending of tube-side pressure drop for Run # 3.

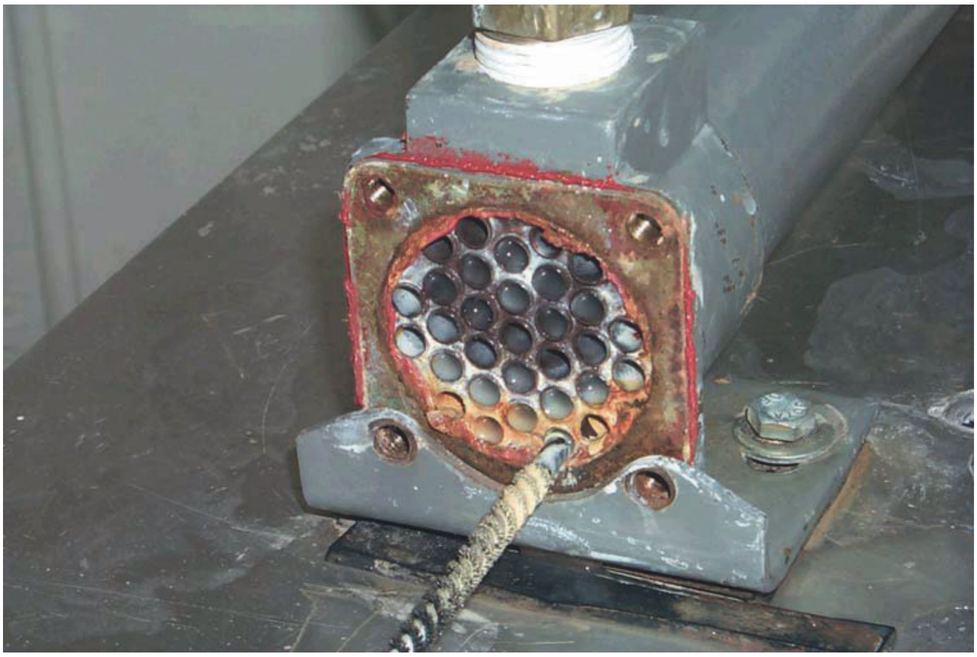


Figure 6. Photograph of the HX tubing with the cleaning brush at the end of the fouling experiment.

The GMDH is an *inductive, self-organizing algebraic model* development approach that creates an *optimal* nonlinear representation of one variable as a function of other related variables. The learning phase may use actual operational data, physics model simulated information, or a combination of the two. The goal is to model a variable *y* as a function of the input set $\{x_1, x_2, \ldots, x_n\}$. At each level of the model structure two of the n-signals, $\{x_i, x_j\}$ are used to generate a second order polynomial model of the form

$$y = A + Bx_i + Cx_j + Dx_i^2 + Ex_j^2 + Fx_i x_j \quad (7)$$

This procedure is repeated for all pairs $\{x_i, x_j\}$. The prediction of y at each layer will improve until the prediction error is within desired limits. The general approximation is called the Kolmogorov-Gabor polynomial and is given by Equation (8). The model-building algorithm is terminated at this point.

$$y = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{ijk} x_i x_j x_k + \ldots \quad (8)$$

$\{a_0, a_i, a_{ij}, \ldots\}$ is the vector of model coefficients, which are estimated by the GMDH algorithm. The details of the algorithm and its implementation are given in Ref. [10].

5.2 Tracking Heat Exchanger Performance

The first step in implementing the GMDH for process variable prediction is to develop a normal operation model. The model structure is defined in Section 5.1. The ideal situation is to develop the model using operational database. In its absence, we suggest the use of high-fidelity physics models to generate the necessary information. In some cases operational data and physics-derived information may be used to generalize the empirical model for a wide range of situations.

In this study, a GMDH model was first developed to characterize the dynamics of the HX cold-side outlet temperature, and then used for the prediction of this variable during future operation. Inlet temperatures on the hot and cold sides of the HX were used as inputs. Figure 7 shows a comparison of the measured and GMDH-predicted values of this temperature with fouling (Run # 2). The residuals between the measured and predicted values are plotted in Figure 8. The overall time behaviour of the cold-side outlet temperature prediction error and the percent change in the thermal resistance are similar. This is an important observation that is useful in isolating the fault type. When the degradation is dominated by the fouling phenomenon, the thermal resistance and model prediction error of the outlet temperature have similar trends, eventually attaining asymptotic values.

6 CONCLUDING REMARKS

Continuous monitoring of process units in chemical industry, power industry, and other process industries is critical for economic growth and environmental protection. Early detection of anomalies in process units must be made part of the overall monitoring system and for maintenance planning, repair, replacement, and resource allocation. The results of the research presented in this paper demonstrate the successful implementation of empirical models developed from operational measurements. The technology is sufficiently advanced to the point where the implementation may be made part of the standard procedure and be incorporated into a computerized maintenance management system (CMMS). Work is currently underway in applying this technique to process units in a coal-fired power plant, with the focus on the coal pulverizer system.

7 REFERENCES

[1] Upadhyaya BR & Huang X. (2004) *Degradation Monitoring and Fault Diagnosis of Heat Exchanger Systems,* Research Report prepared for Emerson Process Management, EPMPT-Annual-04, University of Tennessee.

[2] Melo LF et al. (1988) Particle Transport in Fouling Caused by KAOLIN-Water Suspensions on Copper Tubes, *Canadian Journal of Chemical Engineering,* 66.

[3] Middis J et al. (1990) Particulate Fouling in Heat Exchangers with Enhanced Surfaces, *18th Australasian Chemical Engineering Conference,* Auckland, New Zealand.

[4] Chamra LM et al. (1993) Effect of Particle Size and Size Distribution on Particulate Fouling in Enhanced Tubes, *Enhanced Heat Transfer,* 1(1).

[5] Muller-Steinhagen H et al. (1988) Particulate Fouling in Heat Exchangers, *Transactions of EMCh,* 15(3).

[6] Farlow SJ. (1984) Self-organizing Methods in Modeling: GMDH-type Algorithms, Marcel Dekker, New York.

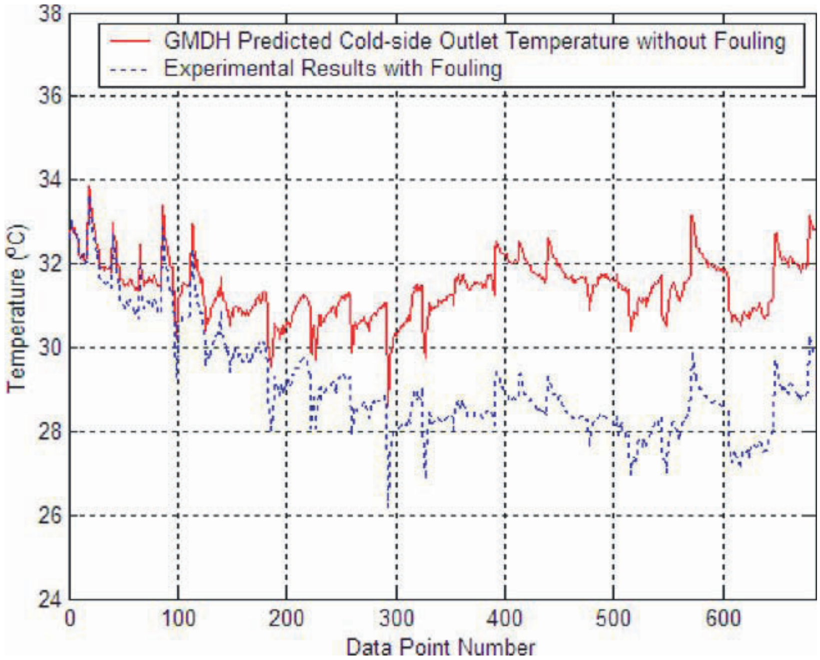


Figure 7. Comparison of the measured HX cold-side (shell side) outlet temperature and the prediction using the GMDH model.

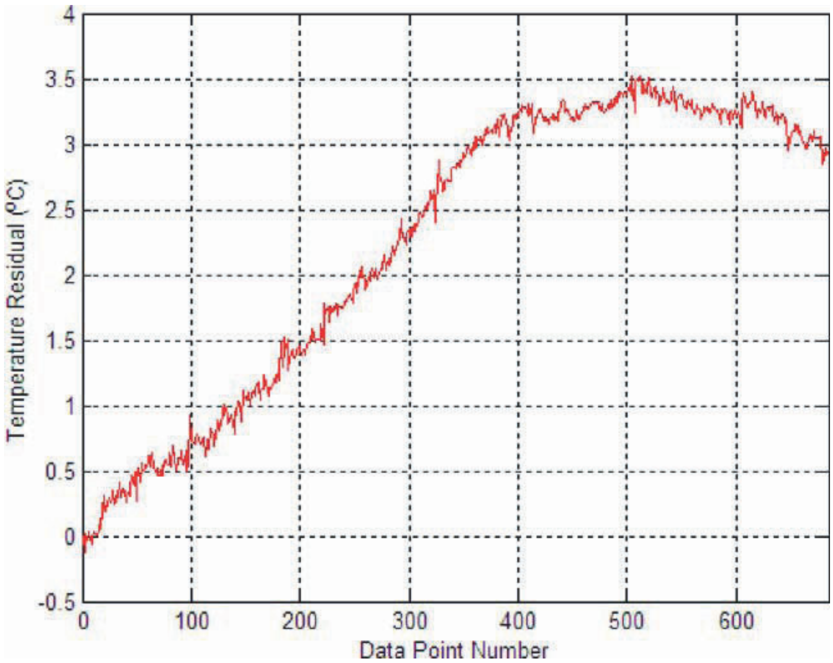


Figure 8. Trending of the error between the measured cold-side outlet temperature and its predicted value from the GMDH model.

[7] Upadhyaya BR et al. (2005) Data-Based Monitoring and Experimental Studies of Particulate Fouling in Industrial Heat Exchangers, *ECI Conference on Heat Exchanger Fouling and Cleaning – Challenges and Opportunities,* Kloster Irsee, Germany.

[8] Goncalves IM, Ting DKS, Ferreira PB & Upadhyaya BR. (2005) Monitoring an Experimental Reactor Using the Group Method of Data Handling Approach, *Nuclear Technology,* 149 (1), pp. 101-109.

[9] Lu B & Upadhyaya BR. (2005) Monitoring and Fault Diagnosis of the Steam Generator System of a Nuclear Power Plant Using Data-Driven Modeling and Residual Space Analysis, *Annals of Nuclear Energy,* 32, pp. 897-912.

[10] Ferreira PB & Upadhyaya BR. (1999) *Incipient Fault Detection and Isolation of Sensors and Field Devices,* Research

[11] Report, The University of Tennessee, UTNE/BRU/99-02.

8 ACKNOWLEDGMENTS

This work was performed under a research contract funded by Emerson Process Management, Eden Prairie, Minnesota, USA, with The University of Tennessee, Knoxville.

DEVELOPMENT OF 3-D SOUND LOCALIZATION SYSTEM BY BINAURAL MODEL

Satoshi Horihata [1] , Zhong Zhang [2], Tetsuo Miyake [2], and Takashi Imamura [2]

[1] *Laboratory of Physics, School of Dentistry at Matsudo, Nihon University, Sakaecho-nishi, Mtsudo, Chiba 271-8587 Japan.*

[2] *Department of Production Systems Engineering, Toyohashi University of Technology, Hibarigaoka, Tempaku-cho, Toyohashi 441-8580, Japan.*

Abstract: In this research, the sound source direction is decided by two angles, one is the horizontal angle α, and the other is vertical angle β. The former is presumed from the difference of the time of arrival of the sound from the source to a right and left ear, and the latter is presumed from spectrum correlation analysis using HRTF(Head Related Transfer Function). HRTF is a transfer function of spectrum modulation caused because the sound is diffracted and reflects by the ruggedness of the head and the earlobe. HRTF was measured beforehand for all directions by using a white noise in preparation for the analysis, and the HRTF database was made. Additionally, the information of the sound for the localization is prepared as knowledge under assumption of having heard it. When the sound is observed, the directional information of the sound is taken out by using information that system has as knowledge. The direction location experiment was done in the anechoic chamber for some sources, therefore the correct presumption rate became about 90% on the average, and was able to show the possibility of the sound localization by this technique.

Key Words: Sound Localization, Binaural Model, Transfer Function, Head Related Transfer Function

1 INTRODUCTION

The robot (electric machine and machine) has evolved to meet various needs in the industrial world and the home flexibly today. However, most of them are still moves by a one-sided instruction of man. In the future, the robot that will think for myself, judges, and acts is needed. It is demanded for the robot to be able to recognize surrounding circumstances by oneself, and to decide the action that should be taken based on the information. It is thought that corresponds to man's sense, sight, aural, sense of smell, and the function of the sense of touch etc. are necessary to recognize surrounding circumstances. In aural, when man listens to a certain sound, the direction of coming, the distance, and the kind of sound can be roughly presumed. It is said, "Sound Localization" to presume the direction of coming among these. The researches on the sound localization were advanced than before by many fields. However, it is usually a technique that uses a lot of microphone arrays. Moreover, the one to need additional movement to rotate for them even if it's with two microphones. It aims at the achievement of three dimension sound localization with two fixed microphones that assume the application of the humanoid robot to the auditory system in this research. Then, binaural model using dummy head (model with built-in mike that imitated man's head) was made as man's auditory model.

In this research, the sound source direction is decided by two angles, one is the horizontal angle α, and the other is vertical angle β. The former is presumed from the difference of the time of arrival of the sound from the source to a right and left ear, and the latter is presumed from spectrum correlation analysis using HRTF (Head Related Transfer Function). HRTF is a transfer function of the sound that changes depending on the direction of coming, and spectrum modulation caused because the sound is diffracted and reflects by the ruggedness of the head and the earlobe. As the mechanism of the human body has the element of uncertainty, it is difficult to evaluate human senses. The observation of human's information processing is indispensable to do the study of human senses. Our research of sound localization is one in aural mechanism. This is the signal processing system by the pinna model which uses two microphones to apply to the auditory system of robots.

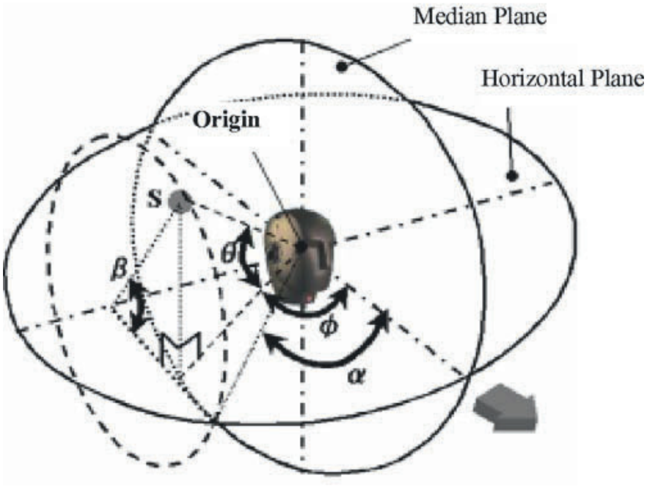
Our researches are shown in the following.

2 3-D SOUND LOCALIZATION USING TRANSFER FUNCTION OF THE BINAURAL MODEL

When we hear a sound, we are able to guess the direction and the position of the sound source. This ability we have is called "sound localization". If this ability is mechanically engineered, this can be applied to aural of robots. And it is possible to use this system as assistance of the physically handicapped person. The Principle of sound direction localizations is shown below.

Up to this time, only some studies about sound localization have been done. In these studies, the array of 3 or more microphones has usually been used, but is not suitable to apply to the auditory system of robots. On the other hand, using 2 microphones is not general and is limited in robotic usage, but is a better option. The purpose of this study is the realization of sound localization using two microphones. The ear model is designed to have different acoustical characteristics (HRTF; Head Related Transfer Function) in every direction. We measured the HRTF in each direction, and have made a database about it. If we can select the direction of the observed sound by using signals of each microphone from the HRTF database, we will be able to know the direction of each sound source. The practicality of this system is being examined by sound localization experimental apparatus including a dummy head. The coordinates of sound source location are decided by an azimuthal angle φ and elevation angle θ. In our system, the horizontal angle α and the vertical angle β are used. The relation between (φ, θ) and (α, β) is shown by Fig.1 and the following equation.

$$\varphi = \tan^{-1}\left(\frac{\tan\alpha}{\cos\beta}\right) \quad , \quad \theta = \cos^{-1}\left(\frac{\sin\alpha}{\sin\varphi}\right) \tag{1}$$



φ: Azimuthal angle, θ: Elevation angle,

α: Horizontal angle, β: Vertical angle

Fig.1 Direction of sound.

The horizontal angle α can be presumed by calculating the time that the sound travel from the sound source to the left and right microphones. This is called Binaural Time Difference (BTD). The vertical angle β is presumed by frequency response of HRTF. In this study, we've made the HRTF database in the anechoic room using white noise practically. We assume that the observed sound has been heard before. And the sound is recorded on a data base as knowledge. In an analysis, α is calculated from BTD of the observation signal, and a group which should refer to the HRTF data base is decided. Next, ratio, *Rhr*, of the standard HRTF and referred HRTF are calculated respectively. On the other hand, ratio, *Rsr*, of an observed signal's spectrum and the standard spectrum is calculated. And, the correlation coefficient γ of *Rhr* and *Rsr* is calculated by the following equation.

$$\gamma = \frac{\int_{-\infty}^{\infty}(Rsr(f)-\overline{Rsr}))(Rhr(f)-\overline{Rhr})df}{\sqrt{\int_{-\infty}^{\infty}(Rsr(f)-\overline{Rsr})^2 df}\sqrt{\int_{-\infty}^{\infty}(Rhr(f)-\overline{Rhr})^2 df}} \tag{2}$$

From the maximum value of γ, the vertical angle β is presumed.

Table.1 Sound source direction of HRTF data base

		[deg]
[deg]	0 20	0, 30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 330
	30 40	0, 36, 72, 108, 144, 180, 216, 252, 288, 324
	50	0, 45, 90, 135, 180, 225, 270, 315
	60	0, 60, 120, 180, 240, 300
	70 80	0, 90, 180, 270

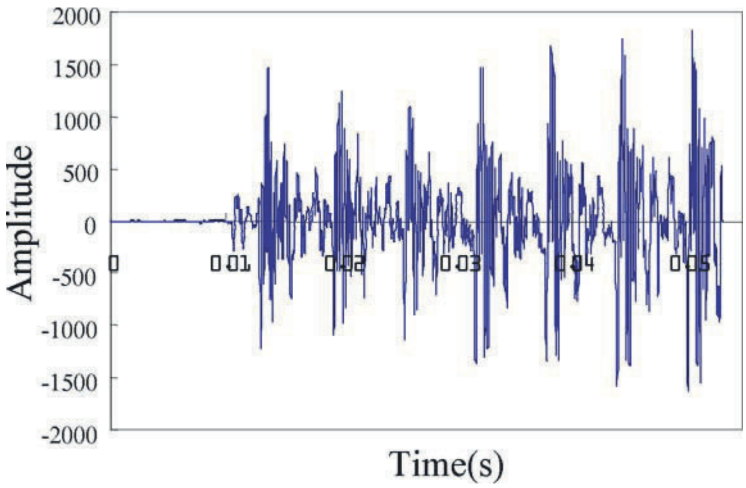


Fig.2 Wave form of vowel /a/ of Japan.

Then we did experiments of sound localization using a man's voice /a/ as the sound source. About experimental conditions, the distance from the starting point to the sound source is 2m and the sampling frequency is 25.6 kHz. The distance between the dummy head's both ears is 14cm. The angle α at the interval of 10°is measured from 0°to 80°. Figure shows the interval of β corresponding to each α. The total of the observation points is 78 points. Table 1 shows observation points of HRTF data base. In this research, Japanese vowels are used as a sound source. Here, we show the analytical result of vowel /a/ of Japan. The wave form of vowel /a/ of Japanese is shown in Fig.2.

3 ANALYTICA RESULTS

3.1 Presumption of horizontal angle

Fig.3 shows the relation between α and BTD obtained by the experiment. Here, α is presumed by a half-adjust at intervals of 10°. In this study, we obtain approximation straight line from the curve that has been obtained because of the experiment in Fig.3. The obtained equation is $\alpha=1.031\times10^5\times$BTD. The correct answers of α are very high in our system.

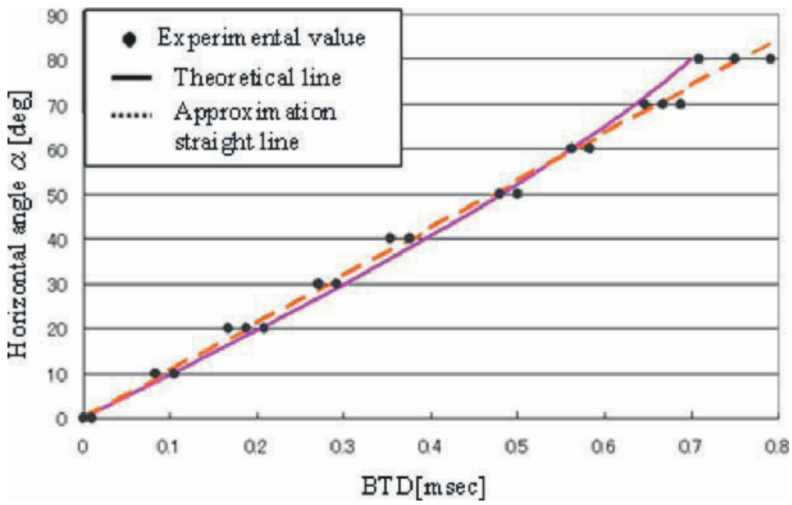


Fig.3α - BTD curve.

3.2 Presumption of vertical corner β

HRTF was measured beforehand for all directions by using a white noise in preparation for the analysis, and the HRTF database was made. Additionally, the information of the sound for the localization is prepared as knowledge under assumption of having heard it. When the sound is observed, the directional information included in the sound is taken out by using information that has it as knowledge. The direction of the source is presumed by collating this HRTF with the database.

Fig.4 shows the presumption result of β obtained by the correlation analysis in the case of α=60° and β=90°. Because the calculation result of the correlation value was the highest in β=90°, it can be said that β will be correctly presumed. Figure 5 shows the correct answer rate of the vertical angle β of the sound of the direction of each α. The range of frequency to obtain correlation value is 234.375-4687.5Hz.

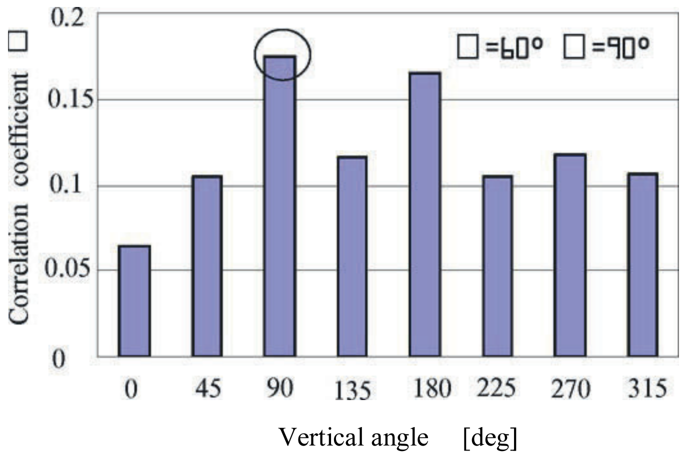
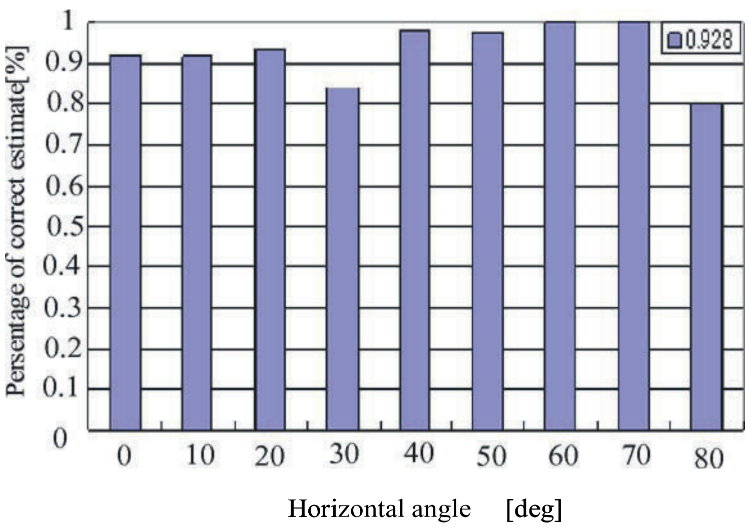


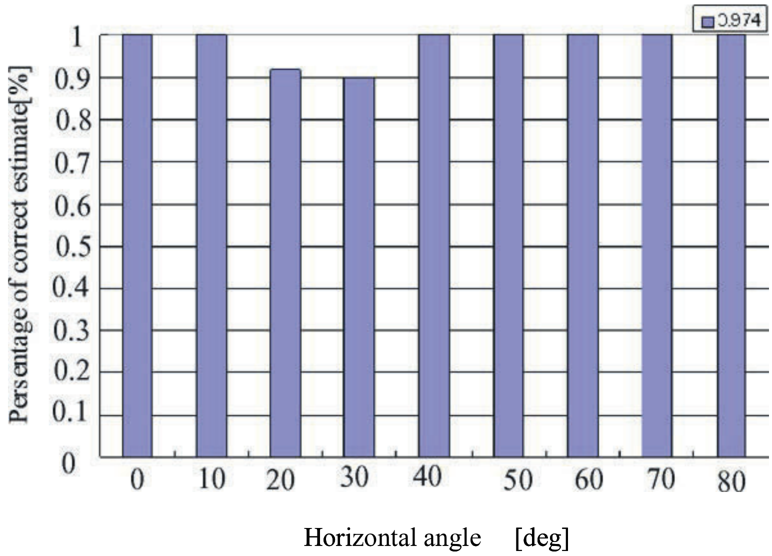
Fig.4 Comparison of correlation coefficient.

Figure 5(a) is an analytical result of the /a/ continuously generated. The average presumption rates obtained about 92%. Now, we are examining the reason why the miss-presumption increases when α is small. As a result, when the sound source was near the median plane of the binaural system, the percentages of correct answers of β were very high. As the sound source was far away from the median plane, the percentages of correct answers of β had decreased. The reason is that the difference of each HRTF data is small. Figure 5(b) is an analytical result of the /a/ intermittently generated. The average presumption rates obtained about 97%. From this result, it has been understood that intermittent sounds are more suitable for the sound direction localization than continuous sounds. Because information on the sound source increased, the accuracy of the location is assumed to be improved.

We were able to show the possibility of the sound localization by this technique. In the future, the three dimension sound direction localization system is constructed by the spectrum analysis including voice recognition.



Result of continuously generated /a/.



(b) Result of intermittently generated /a/.

Fig.5 Localization result of each direction.

4 CONCLUSIONS

In this report, we showed the 3-D sound localization method that used transfer function of the binaural model. As a result of the sound direction localization experiment by the dummy head, we were able to show effectiveness of our system. In the future, we want to solve some problems, and to improve our system to a more practicable system.

The direction location experiment was done in the anechoic chamber for some sources; therefore the correct presumption rate became about 90% on the average, and was able to show the possibility of the sound localization by this technique. These accurately locate as a problem in the future about a specific direction because there are a lot of miss-presumptions. Correspondence to another source that builds in sound source recognition and the localization in real time, etc. are raised.

5 REFERENCES

[1] Shuushi Makita, Takehiko Katayama, Norihiko Yamada, Satoshi Horihata and Tetsuo Miyake: 3-D Sound Localization Using Head Related Transfer Function, Mechanical Engineering Congress, Proceedings of MECJ-03, Vol. , pp.229-230, 2003, in Japanese.

[2] Takehiko Katayama, Satoshi Horihata and Tetsuo Miyake: 3-D Sound Localization by Binaural Model, The Japan Society of Mechanical Engineers Tokai Branch Congress, Proceedings of JSME Tokai Branch - 04, No.043-1, pp.223-224, 2004, in Japanese.

DESIGNING AVERAGE COMPLEX REAL--SIGNAL MOTHER WAVELET AND APPLYING IT IN ABNORMAL SIGNAL DETECTION

Zhong ZHANG [a], Hiroki IKEUCHI [a], Tetsuo MIYAKE [a], Satoshi HORIHATA [b], Hideaki ISHII [c]

[a] *Department of Production System Engineering, Toyohashi University of Technology, Japan*

[b] *School of Dentistry at Matsudo, Nihon University, Japan.*

[c] *JTEKT Corporation*

Abstract: In this study, in order to improve the problem of abnormal signal selection, the technique of constituting an Average Real signal Mother Wavelet (A-RMW) from two or more real signals is proposed and its effectiveness is confirmed. The main results obtained can be summarized as follows: (1) For constructing A-RMW from two or more signals, the Symmetric Complex Real signal Mother Wavelet (SCRMW) first was constructed for each signal and then the average SCRMW was obtained by adding them. Finally the average SCRMW obtained above was normalized and the A-RMW was obtained. (2) The proposed construction method of the SCRMW is different to that of the CRMW [9] in that the phase information of all the frequency components has been cut off and symmetry can be obtained.

Key Words: Wavelet transform, Wavelet, Real signal mother wavelet, Abnormal signal detection

1 INTRODUCTION

It is well known that the Short Time Fourier Transform (STFT) and the Wavelet Transform (WT) are used widely as time-frequency analysis methods since they express the features of an unsteady signal clearly on the time and frequency plane [1]. However the STFT has fixed time resolution and frequency resolution at all frequencies and the determination of window width is difficult. Compared to this, the WT has a suitable resolution at each frequency because it carries out expansion reduction of the function called the Mother Wavelet (MW) and changes the time and frequency resolution according to frequency. Therefore, the WT has been applied to signal processing in various fields, and has received attention [2, 3].

For the MT, all functions can be used as the MW if they satisfy the admissibility condition and many types of MWs, for example, the complex type and real type and so on have been proposed [4]. Understandably, analysis results obtained using different MWs are different. So in signal analysis, one often has to first examine the characteristics of the MWs, and then choose one. For example, when analyzing an unfamiliar signal, the Gabor function [6] and the RI-Spline wavelet [5] generally produce better results than other wavelets since they have good localization in the frequency and time domain, and easily matche the frequency and scale of the signal. However, for detecting abnormal signals which appear unsteady, the Gabor function and the RI-Spline wavelet cannot always give satisfactory results.

In addition, the WT can be considered a kind of similarity transform [7, 8]. A signal $f(t)$ is measured by using the wavelet functions $\psi_{a,b}(t)$ that are defined from the MW $\psi(t)$ as a scale, and the similarity between the signal $f(t)$ and wavelet functions $\psi_{a,b}(t)$ is evaluated by the coefficient of the WT. The authors [7] took advantage of this feature and constituted a Real signal Mother Wavelet (RMW) from an abnormal signal. Then the WT was carried out by using the RMW and the abnormal signal was detected by evaluating the correlation between the RMW and the abnormal signal using the wavelet coefficient. This method was very useful for detecting the abnormal signal although it had the following two problems: 1) The signal's phase information could not be obtained since the RMW constructed was a real type MW and the wavelet coefficients oscillated. 2) The RMW is difficult to form when it is not clear what portion of the signal is suitable for the RMW. In this case, it was difficult to narrow down to one with the WT using the Gabor function and the RI-Spline wavelet.

In order to suppress oscillation of the analysis results, the authors proposed a Complex Real signal Mother Wavelet (CRMW) [9] and confirmed its effectiveness by applying it to abnormal vibration detection. In this study, in order to ease the difficult problem of the signal selection, the technique of constituting an average RMW from two or more real signals is proposed. The important technique of this method is the construction of a Symmetric Complex Real signal Mother Wavelet (SCRMW), in which the phase information of all the frequency components has been cut off. Furthermore, the proposal method is applied to the abnormal sound source detection of a car, and its effectiveness is examined.

2 RMW AND WAVELET INSTANTANEOUS CORRELATION

The WT transforms the signal $f(t)$ which has a finite energy into a function which has two variables, frequency ($1/a$, a : scale) and time b . The WT is defined by the following formula.

$$W(a,b) = \int_{-\infty}^{\infty} f(t)\overline{\psi_{a,b}(t)}dt, \quad (1)$$

$$\psi_{a,b}(t) = a^{-1/2}\psi(\frac{t-b}{a}),$$

where $\overline{\psi_{a,b}(t)}$ and $\psi_{a,b}(t)$ are conjugate complex functions of each other. Here, the function $\psi(t)$ is the MW which satisfies the following admissibility condition.

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty. \quad (2)$$

Here, $\hat{\psi}(\omega)$ means the Fourier transform of $\psi(t)$, ω denotes the angular frequency ($\omega = 2\pi f$), and f the frequency. This condition can be simplified to the next equation when $\psi(t)$ tends to zero as t approaches infinity.

$$\int_{-\infty}^{\infty} \psi(t)dt = 0. \quad (3)$$

Formula (1) shows that the wavelet transform achieves the time-frequency analysis by transforming the signal $f(t)$ into the function $W(a,b)$ which has two variables, frequency ($1/a$) and time b . Generally, the abnormal signal consists of many components, its strength is various and its generating time is irregular. The WT can show these features clearly in the time-frequency plane, but it cannot detect and evaluate its features at the same time because the common MW performs band pass filtering. Therefore, creating a technique for the detection and evaluation of abnormal signals is still a very difficult task.

Moreover, as is shown in Eq. (3), all functions can be used as the MW if they are functions such that their average value is zero and the amplitude becomes zero sufficiently quickly at a distant point. Therefore, the real RMW can be constructed by multiplying the real signal with a window function and removing the average which makes it approach zero sufficiently quickly at a distant point [7]. Furthermore, the authors proposed the method for constructing a CRMW using the frequency characteristic of the complex wavelet function [9] and defined the Wavelet Instantaneous Correlation Value (WICV) $R(b)$, which is a value $|w(a,b)|$ obtained by the CRMW in scale a =1 was shown as follows:

$$R(b) = |w(a=1,b)|. \quad (4)$$

By using $R(b)$, the detection and evaluation of the abnormal signal's strength and generating time can be achieved at the same time.

The procedure for constructing the CRMW can be summarized as follows: 1) First the characteristic portion of the real signal is isolated and then the real signal is multiplied with a window function. Next the average is removed thereby making it approach zero sufficiently quickly at a distant point and obtaining the RMW $\psi_R(t)$. 2) The RMW is normalized so that the norm $\|\psi_R\|$ =1.

$$\|\psi_R\| = \left[\int_{-\infty}^{\infty} \psi_R(t)^2 dt\right]^{1/2} = 1. \quad (5)$$

3) The Fourier transform of the RMW $\psi_R(t)$ is performed and its frequency spectrum $\hat{\psi}_R(f)$ is obtained. 4) In the negative frequency domain, $\hat{\psi}_R(f)$ is set to 0, in the positive frequency domain, $\hat{\psi}_R(f)$ is set to $2\hat{\psi}_R(f)$ and the reverse Fourier transform is performed. Finally the CRMW $\psi(t)$ = $\psi_R(t) + i\psi_I(t)$ can then be obtained.

Figure 1(a) shows an example of the RMW $\psi_R(t)$, which was constructed by a model signal, where the Hanning window function has been used. The model signal consists of three sine waves with 400Hz, 800Hz and 1600Hz which can be shown as follows:

$$f(t) = \sin(2\pi 400t) + 0.7\sin(2\pi 800t) + 0.7\sin(2\pi 1600t) \qquad (6)$$

Where t denotes time. Figure 1(b) shows an example of the CRMW $\psi(t)$ that was constructed by the method shown above using the RMW $\psi_R(t)$ shown in Fig. 1(a) and Fig. 1(c) shows their power spectrum $E(f)$. Moreover, the norm of the real component and the imaginary component of the CRMW are set to 1, respectively. As is shown in Fig. 1(b), construction of the CRMW was successful.

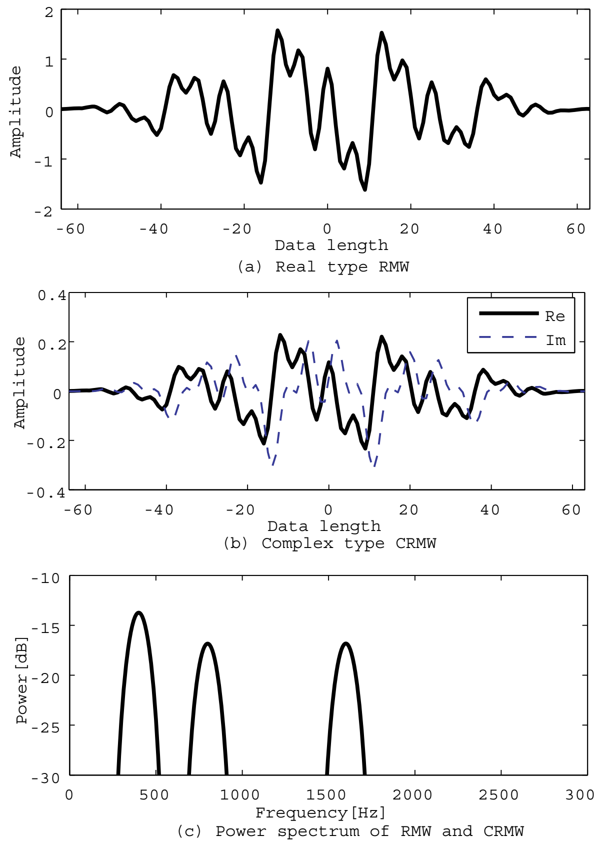


Figure 1 Example of the RMW and CRMW constructed by the model signal.

3 CONSTRUCTION OF AVERAGE REAL SIGNAL MOTHER WAVELET (A-RMW)

When constructing the RMW, there is a difficult problem in that one can not always narrow down the location of an abnormal signal if it is not clear what portion of the signal is suitable for the RMW. Therefore, it is possible that if the selection of the portion of an abnormal signal is not correct, then the abnormal signals can not be detected correctly by using the RMW constructed above. In this case, it was thought that if the RMW was first constructed from multiple locations of the abnormal signal (or multiple abnormal signals) and then the abnormal signal could be detected correctly. Therefore, the construction method shown below of the A-RMW was proposed and the construction technique of the Symmetric Complex Real signal Mother Wavelet (SCRMW) plays an important role in the A-RMW construction.

3.1 Construction method of the SCRMW

Figure 2 shows the flow chart of the construction method of the SCRMW. In the figure, steps 1 to 4 are the same as the process of CRMW [9]. However step 5 is different from the CRMW in that the real part of $\hat{\psi}(f)$ is set to $\sqrt{(\hat{\psi}_R(f))^2 + (\hat{\psi}_I(f))^2}$, and the imaginary part is set zero. As a result, the phase information of all the frequency components has been cut off. In step 6, the inverse Fourier transform was carried out and the SCRMW $\psi(t) = \psi_R(t) + i\psi_I(t)$ can be obtained.

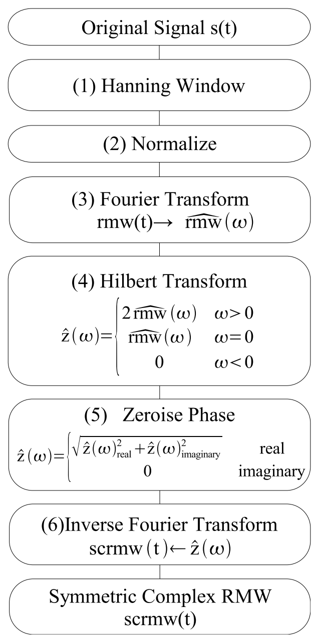


Figure 2. The method of constructing the SCRMW

Figure 3(a) shows the SCRMW constructed using the model signal shown in Eq. (6) and (b) shows the CRMW constructed using the same model signal for comparison. From Fig. 3(a) it is clear that the energy of the SCRMW is concentrated around the center of the SCRMW. In addition, the real component of the SCRMW has symmetry and the imaginary component has anti-symmetry. Compared to the SCRMW, the CRMW holds the original signal's phase information and does not have symmetry.

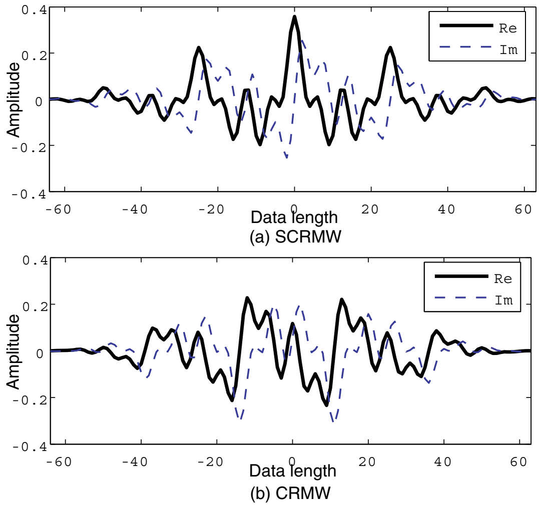


Figure 3. Example of the SCRMW and the CRMW

3.2 Construction method of the A-RMW

For construction of the A-RMW, the method proposed by authors, which connecting multiple signals in the time domain was used [10]. Figure 4(a) shows the average RMW obtained by connecting two signals. In this figure, we use two model signals, one with the components of 400, 800, and 1600Hz shown in Eq. 6 and another one with the components of 300, 600, and 1200Hz shown in the next equation.

$$f(k) = \sin(2\pi 300t) + 0.7\sin(2\pi 900t) + 0.7\sin(2\pi 1200t) \tag{7}$$

In the construction of the average CMRW, first two data sets with 128 points in which one was extracted from the signal shown in Eq. (6) and another one was extracted from the signal shown in Eq. (7), respectively, were arranged continuously and a synthetic signal shown in Fig. 4(a) was obtained. Then multiplying the synthetic signal with the window function, and removing the average yielded the average RMW. Figure 4(b) shows an example of the average RMW constructed. An average CRMW finally was obtained by the construction method of the CRMW using the average RMW as the original signal and is shown in Figure 4(c). As is shown in Fig. 4(c), the average characteristic of the two signals can not be obtained from the average CMRW since using the construction method of CMRW, the phase information of the two signals were held and a discontinuity arose in the connection portion.

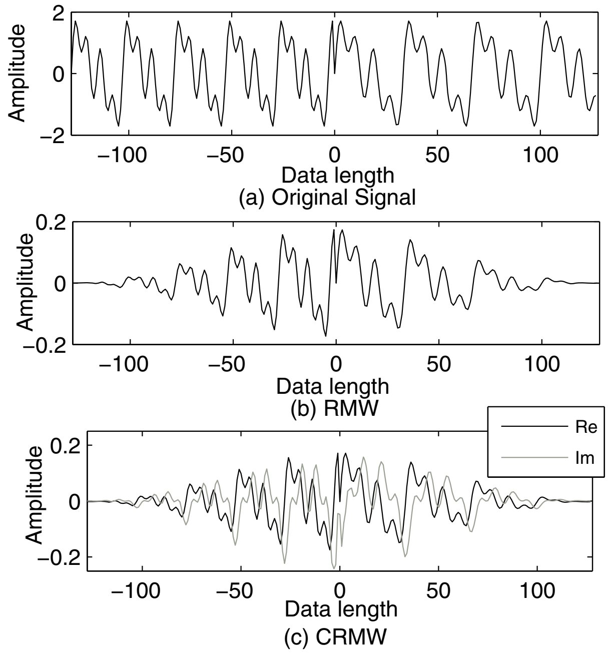


Figure 4. Example of the average CRMW

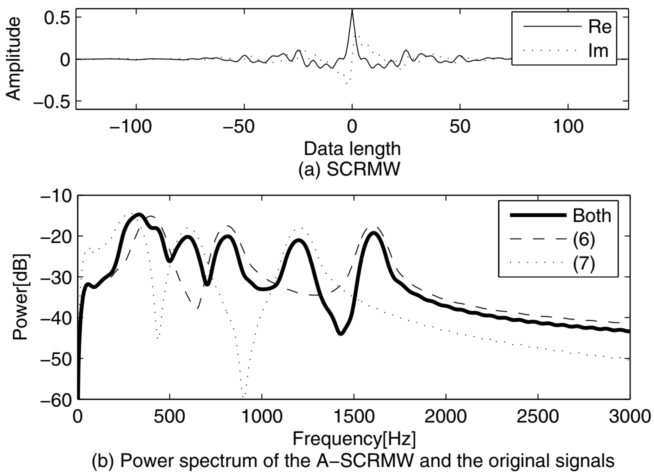


Figure 5. Example of the Average SCRMW

In order to improve the discontinuity problem in the connection portion, the constructing method of the SCRMW was used instead of the construction method of the CRMW and the results are shown in Fig. 5. Here, Fig. 5(a) shows the average SCRMW and (b) its power spectrum. As is shown in Fig. 5, the average SCRMW has the SCRMW's property that the real component has symmetry and the imaginary component is anti-symmetric, and the wave of an average SCRMW becomes smoother than an average CRMW although the frequency property which high frequency components have been increased is not good because the average RMW shown in Fig.4 (a) has a discontinuous part in the connection portion.

3.3 Problems and improving method

Regarding the average SCRMW shown above, the following problems can be observed. 1) The data length of an average SCRMW becomes longer than the data quarried out. 2) The average SCRMW's frequency property is not good because the discontinuity is in the connection part of two (or more) signals. Especially for problem 2, improvement is very difficult because the original data of the average RMW shown in Fig.4 (b) is discontinuous. In order to overcome the above drawbacks, a new construction method that uses the SCRMW's features was developed and it is shown in Fig. 6. That is, for two or more signals, the SCRMW first were constructed for each signal and then the average SCRMW was obtained by adding them. Finally the average SCRMW obtained above was normalized and the A-RMW was obtained. Figure 7 shows an example of the A-RMW, where (a) is an example of the A-RMW constituted, and (b) is its frequency characteristic. By comparing Fig. 7 and Fig. 5, it is clear that in the case of the average SCRMW shown in Fig. 5, it has some frequency components which the original model signal does not have. However, in the case of the A-RMW, this phenomenon has been improved.

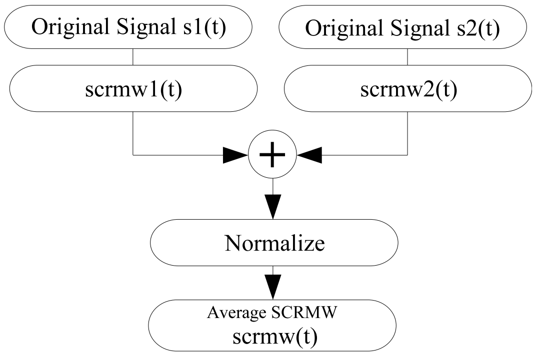


Figure 6. Method of generating the A-RMW

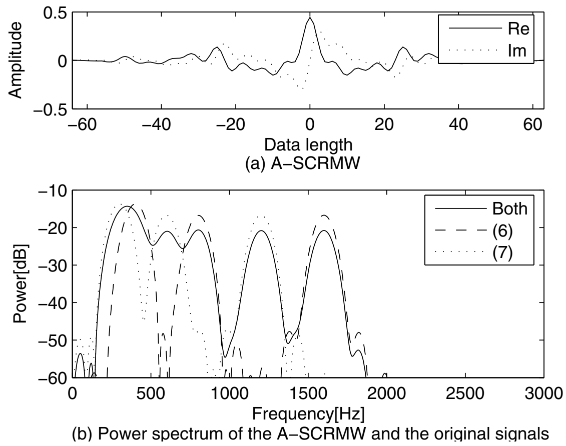


Figure 7. Example of the A-RMW constructed from the SCRMW

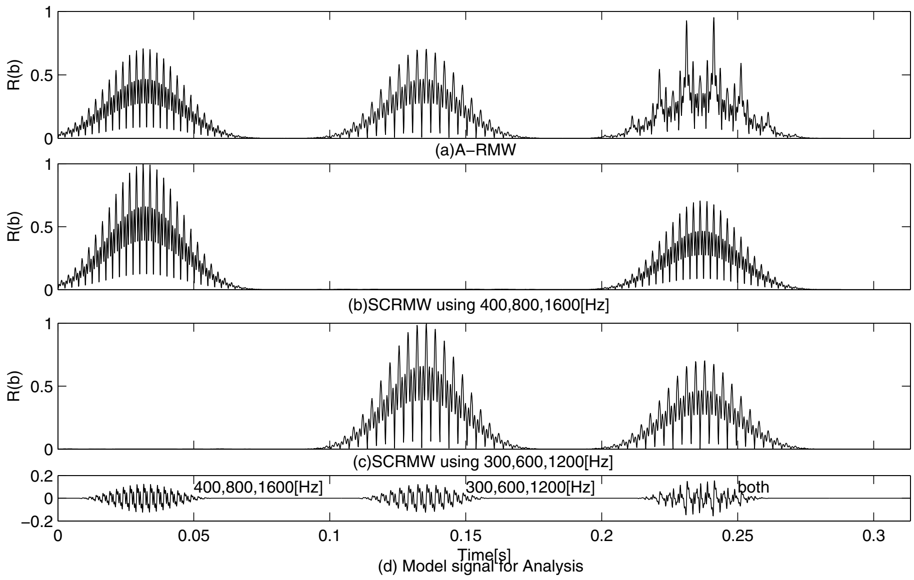


Figure 8. Example of analysis results by using the A-RMWs

3.4 Test of the A-RMW by Simulation

To confirm the features of the A-RMW, a simulation using the model signals was carried out. Figure 8 shows a model signal for analysis and the results obtained by the WICV. The model signal shown in (d) has three parts, which can be shown in Eq. (6), Eq. (7) and Eq. (6) + Eq. (7), respectively. The simulation results obtained by using the A-RMW and the SCRMW are shown in Fig. 8(a), (b) and (c), where (a) shows the results obtained by the A-RMW constructed using the model signal Eq. (6)+Eq. (7), (b) obtained by the SCRMW using the model signal Eq. (6) and (c) obtained by the SCRMW using the model signal Eq. (7). Furthermore, in order to show the results clearly, the length of those A-RMW and SCRMW are 512 points. As is shown in Fig. 8, in the case of (b), the signal of Eq. (7) was not detected and in the case of (c), the signal of Eq. (6) was not detected. Compared to this, in the case of (a), both signals have been detected. Therefore, we can conclude that the construction method of the A-RMW shown in Fig. 6 is successful.

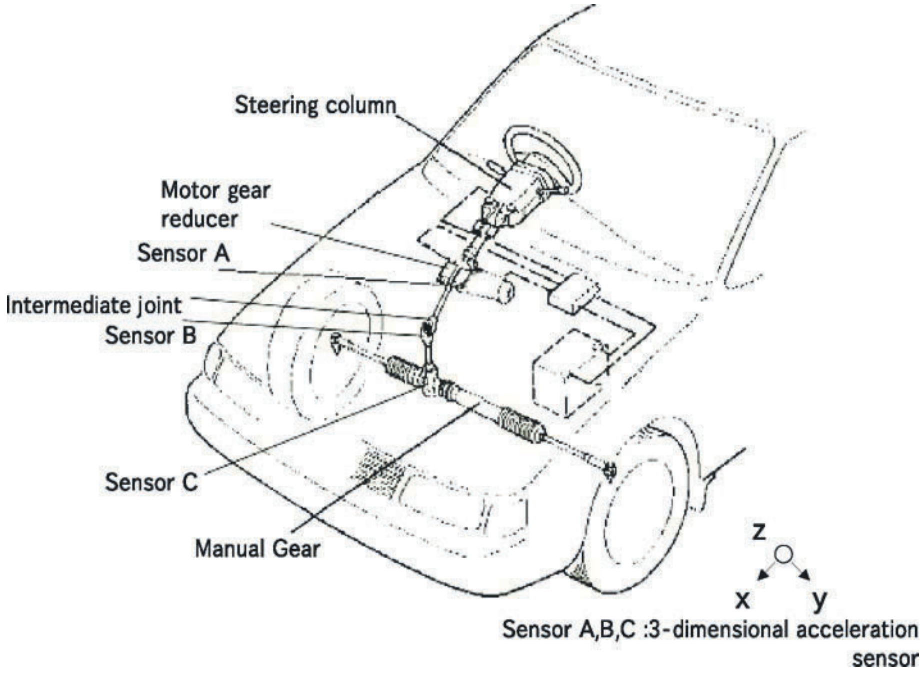


Figure 9. The position of acceleration sensors

4 APPLICATION ON RATTLE NOISE SOURCE DETECTION

Recently, an Electric Power Steering (EPS), in which the motor is driven only when the steering wheel is turned, has been increasingly adopted in the cars, because it greatly improves fuel performance over conventional hydraulic power steering. However, a rattle noise and a joint noise are occurred from front of vehicle while driving over rough roads when vehicle equipped with the EPS system and these noises become problem now. The EPS has the vibration sources such as motor gear reducer, manual gear, intermediate joint, and so on. In order to improve the rattle noises, it is very important to evaluate rattle noises and identify the correct noise source that occurred by the vibration in the EPS, although it is very difficult since the vehicle interior noise generated by the engine, suspension and tires, does not only by the EPS. Therefore, one of the main problems related to identifying the rattle noises in EPS is how do detect the rattle noises from the many kinds of noises. In this section, we will discuss the rattle noises detection by using the A-RMW.

4.1 A-RMW Constructed by Real Sound Signals

Figure 9 shows the position in which the acceleration sensors were installed. As is shown in Fig. 9, three 3-dimension acceleration sensors were installed in each position. A microphone was set up under the steering wheel. The experiments were carried out when a car was driven on a road, and nine vibration data (three sensors and three directions) and a sound data were obtained.

Figure 10 shows example of the sound signals and the A-RMW constructed, where (a) and (b) are the sound signals, (c) is the A-RMW constructed from (a) and (b) by extracting the portion of the rattle noise, which shown in square area of the sound signals. Here, notes that the low-pass filters which cutoff frequency 1kHz for the sound signal has been used. Figure 11 shows the averaged spectrum obtained by the sound signals shown in Fig. 10(a) and (b), and the spectrum obtained from A-RMW shown in Fig. 10(c). From Fig. 11 it is clear that the spectrums are almost same although the wave shape of the A-RMW is different from that of the sound signals. That is, the average property of the rattle noise has been obtained by using A-RMW. Furthermore, in the next section, the A-RMW would be used to the sound source detection of the rattle noise.

4.2 Results and Discussion

Figure 12 (a) shows the example of the vibration signal, which obtained by acceleration sensor C and direction Z (Z direction of Manual Gear), (b) shows the example of analysis result obtained by the WICV $R(b)$ using the A-RMW, where the A-RMW was constructed by using eight signals of the rattle noises that extracted from sound signals. As is shown in Fig. 12(b), the bigger values than 2 of the $R(b)$ can be obtained at 0.35ms, 0.4ms, 0.75ms and 0.9ms. It means the vibration at these time has same frequency characteristic as the A-RMW that constructed by the rattle noise.

Figure 13 shows the mean values of the $R(b)$ obtained in the calculation range, where horizontal axes shows the number of samples which are the seeds of the A-RMW and vertical axes shows $R(b)$ value. Here, the rattle noise is correctly detected, so that average value is high. As is shown in Fig. 13, the vibrations obtained from the motor gear reducer in the direction of X and Y have bigger WICV $R(b)$ than that of the other vibration signals, that is, these vibrations have the large contribution to the rattle noise. Furthermore, the values of the $R(b)$ become large as the number of samples of the rattle noise increase when the number of samples are small than 6 and the values of the $R(b)$ are almost same when the number of samples are bigger than 6. It predicts that an optimal number of the samples exist for the A-RMW and it is almost six.

Figure 14 shows the mean values of the $R(b)$ that included an error signal, where the error signal was shown as "sample 0" in the figure. The example of the error signal was shown in Fig. 15. Where Fig. 15(a) shows the original error signal which selected from sound signal and (b) shows its power spectrum. By comparing Fig. 15(b) and Fig. 11, it is clear that the spectrum of the error signal and the rattle noise are difference.

The influence of the error signal decreases when a lot of the samples of the rattle noise were used although the selection of the rattle noise for the A-RMW is mistake.

Therefore, we can conclude that the A-RMW constructed from two or more real signals is successful for improving the difficult problem of the signal selection.

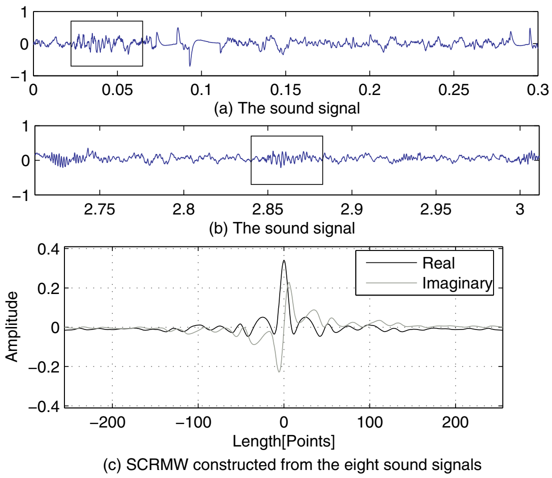


Figure 10. The acoustical signal and A-RMW

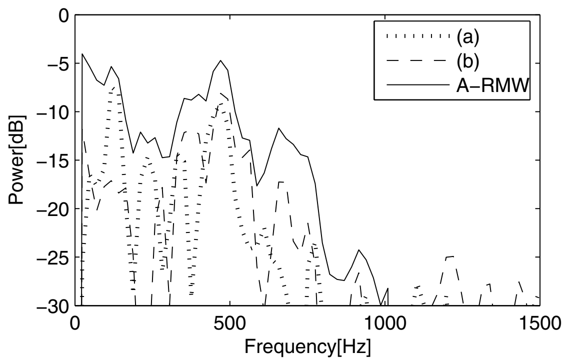


Figure 11. The frequency characteristic of the rattle noise signal and A-RMW

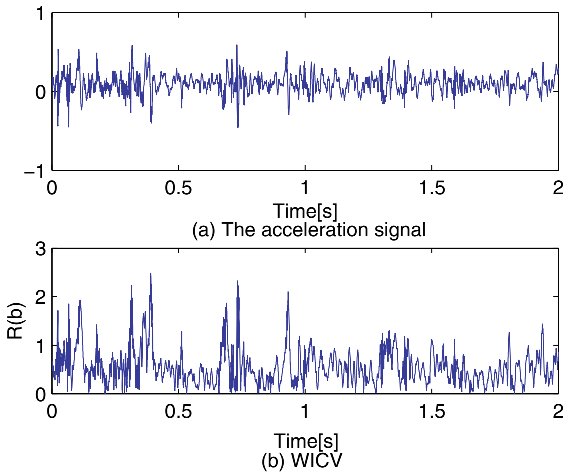


Figure 12. The acceleration signal and analysis result

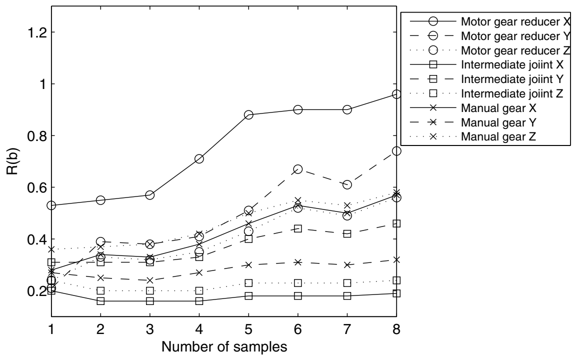


Figure 13. Mean values of WICV $R(b)$

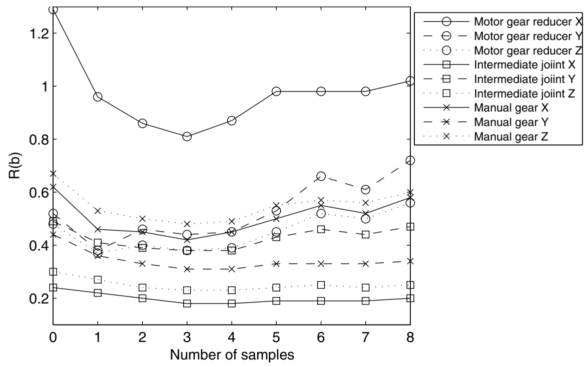


Figure 14. Mean values of WICV (with error signal)

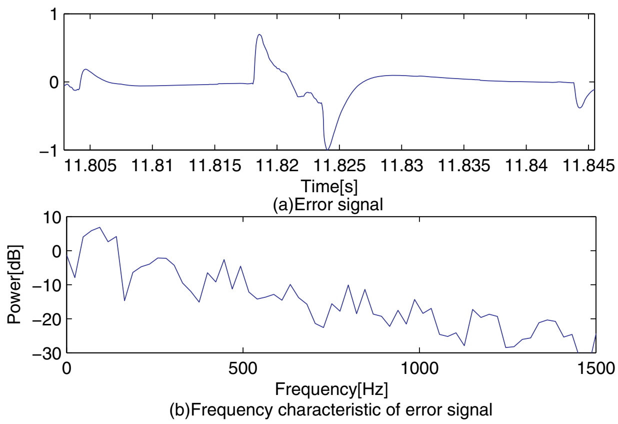


Figure 15. Error signal

5 SUMMARY

In this study, in order to ease the difficult problem of signal selection, the technique of constituting A-RMW from two or more real signals has been proposed and its effectiveness has been confirmed. The main results obtained can be summarized as follows:

1. For constructing the A-RMW from two or more signals, the method in which the SCRMW first was constructed for each signal and then the average SCRMW was obtained by adding them, and finally the average SCRMW obtained was normalized and the A-RMW was obtained was effective.
2. The proposed construction method of the SCRMW is different to the CRMW in that the real part of $\hat{\psi}(f)$ is set to $\sqrt{(\hat{\psi}_R(f))^2 + (\hat{\psi}_I(f))^2}$, and the imaginary part is set to zero. As a result, the phase information of all the frequency components has been cut off and symmetry can be obtained.
3. Furthermore, the proposed method was applied to the rattle noise detection of the EPS, and its effectiveness was examined.

6 REFERENCES

[1] R.L. Allen, D.W. Mills, Signal Analysis, IEEE Press, P.712, 2004.

[2] Z. Zhang, H. Kawabata, and Z. Q. Liu, Electroencephalogram Analysis using Fast Wavelet Transform, International Journal of Computers in Biology and Medicine 31-6, pp.429-440, 2001.

[3] Z. Zhang, H. Kawabata and E. Tomita, A New method of Fractal Analysis Using Wavelet Transform and its Application in Tumbling Flow analysis, JSME International Journal Series B, 46-1, pp.137-144, 2003.

[4] M. Holschneider, Wavelets, An Analysis tools, Oxford Science publications, 1995.

[5] Z. Zhang, H. Kawabata, and Z. Q. Liu, Non-stationary Signal Analysis Using the RI-Spline Wavelet, Integrated Computer-Aided Engineering 8-4, pp.351-362, 2001.

[6] C.K. Chui, An Introduction to Wavelets, Academic Press, 1992.

[7] Z. Zhang and E. Tomita, Knocking Diagnostic using Wavelet Instantaneous Correlation Method, JSAE Review, 23-4, pp.443-449, 2002.

[8] S. Horihata, S. Ishimitsu, T. Miyake, Brain Wave Analysis using Wavelet By Employing Measured Signal as AW, Transactions of the JSME C, 70-694, pp.1795-1801, 2004(in Japanese).

[9] Z. Zhang, S. Horihata, T. Miyake, Complex Real Signal Mother Wavelet and Its Application on Extraction of Abnormal Signals, Proc. of First International Symposium on Advanced Technology of Vibration and Sound (VSTech 2005), pp.319-324, 2005.

[10] H. Ikeuchi, S. Horihata, Z. Zhang and T. Miyake, Detection of Abnormal Signal, by Real Signal Wavelet Transform of EEG, Proc. of the First International Conference on Complex Medical Engineering, pp.671-675, 2005.

FLOW VISUALISATION IN DENSE PHASE PNEUMATIC CONVEYING OF ALUMINA

Cenna AA, Williams KC, Jones MG and Page NW

The University of Newcastle, Australia

Abstract: Pipeline wear is a very complex problem and at present there is limited understanding of the wear mechanisms responsible for the reduction of pipe wall thickness in critical wear areas. The ability to determine the wear mechanisms in these areas holds the key in determining the service life of pneumatic conveying pipelines in industry. In this paper design and construction of a visualisation test rig is presented which enables better understanding of the flow process in pneumatic conveying pipelines. Results from the flow visualisation have been analysed and presented for an insight to the flow patterns in the critical wear areas of pneumatic conveying pipelines.

Wear samples from an industrial pipeline have also been analysed in relation to the flow patterns in the pipelines and bends. The flow patterns observed through the visualization test rig clearly demonstrate its capability to establish the critical wear areas on the pipeline and the bends depending on the pneumatic conveying parameters. Possibility of the development of a realistic predictive model for wear in these critical wear areas have been discussed and presented.

High speed video was used to record the flow patterns inside the bends and immediately after the bends (reacceleration zone). Visualization of flow patterns in the bend clearly showed the complex nature of the flow development in the bends. Impact and sliding of the particle stream on different surfaces clearly showed the critical wear areas in the pipeline and bends. Scanning Electron Microscopy was used to analyse the wear samples from the industrial pipeline. Finally a conceptual model for the wear process is presented.

Key Words: Critical wear, service life, visualisation test, flow patterns, Predictive model, Electron Microscopy, Conceptual model.

1 INTRODUCTION

Pneumatic conveying is a very complex problem of two phase flow and at present there is limited understanding of the flow characteristics in the pipeline, especially through and immediately after the bends. Currently researchers and professionals in the field of pneumatic conveying depend on the concepts and models developed on the basis of steady state conveying conditions. Although these models can be used to estimate the conveying gas pressure and air mass flow rate through the pipeline which provide information on the motive force energy available, the solids (or particle) flow information is limited to a mass flow value prediction which provides no information of the geometric flow structure of the particle stream.

Wear is surface damage that involves progressive material loss due to relative motion between the surface and the contacting substance or substances. Wear poses serious consequences in pneumatic conveying systems with respect to the service life as well as maintenance management. The ability to determine the effects of conveying parameters on wear mechanisms calls for a systematic study of the evolution of the wear profiles on the pipelines due to the particle stream.

In this paper, the design and construction of a visualisation test rig for pneumatic conveying is described. Flow visualization tests were conducted by controlling the solid and air mass flow rates into the system. A high speed video has been used to analyse the flow structure, specifically the flow patterns inside and at the exit of the bends. Using high speed video, the particle stream velocity and associated impact angles in the critical wear area were determined.

2 PNEUMATIC CONVEYING

Pneumatic conveying has gained popularity because it offers a flexible line layout, containment of the conveyed solids and ease of automation. For the case of classification, pneumatic conveying can be categorised into two modes of flow; 1) Dilute phase flow and 2) Dense phase flow. For simplicity, each phase can be categorised in terms of the mass flow ratio (m*), which is defined as the ratio of the mass flow rates of solids to the mass flow rates of conveying air. Thus in terms of the above definition we have (1):

Mode	Mass Flow Ratio
Dilute Phase	0-15
Dense Phase	<15

In dilute phase systems, material is conveyed in suspension in air through the pipeline. Dilute phase (sometimes referred to as lean phase) systems in general employ large volumes of gas at high velocities. The gas stream carries the discrete particles by means of lift and drag forces acting on the individual particles. Dilute phase systems constitute the most widely used of all pneumatic conveying systems. Almost any material can be conveyed in dilute phase, suspension flow through a pipeline, regardless of the particle size, shape or density [2].

In dense phase systems material is in a non-suspension mode in which the particles are generally in contact with each other. Depending upon the major characteristics of the solids and air mass flow rates and the pressure drop through the pipeline, the flow patterns in the dense phase mode can vary from being unstable to stable or an intermediate unstable/stable regime. Figure 1 shows the two modes of dense phase flow diagrammatically as (a) fluidized dense phase or fluidised moving bed type flow and (b) slug or plug flow.

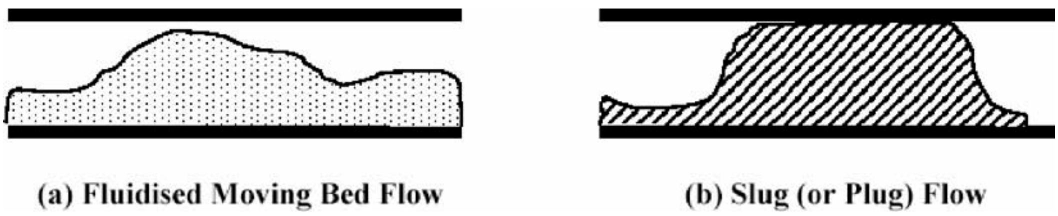


Figure 1: Two Major Modes of Dense Phase

The fluidised dense phase mode is exhibited largely by powdered materials which have good air retention capabilities, whereas, the slug flow mode is generally exhibited by coarse granular materials which have a high level of permeability to the conveying gas. Hence, it is useful to be able to distinguish between those materials capable of dense phase in a conventional pipeline and those restricted to dilute phase only. Furthermore, an ability to identify which mode of dense phase is most suitable for the material is also an advantage.

The boundary between dilute phase flow and dense phase flow is not well defined. For a fixed solids loading ratio, as the superficial gas velocity is increased, the flow condition shifts from a steady dense phase to an unsteady phase and finally to a steady dilute phase as shown in the Zenz diagram (Figure 2) [3]. This shift from stable to unstable transport is marked by a characteristic point called the critical velocity, which has been defined as the transition between dense phase flow and slug flow or plug flow transport. In this paper fluidized dense phase flow conditions for conveying alumina is the focus of the work.

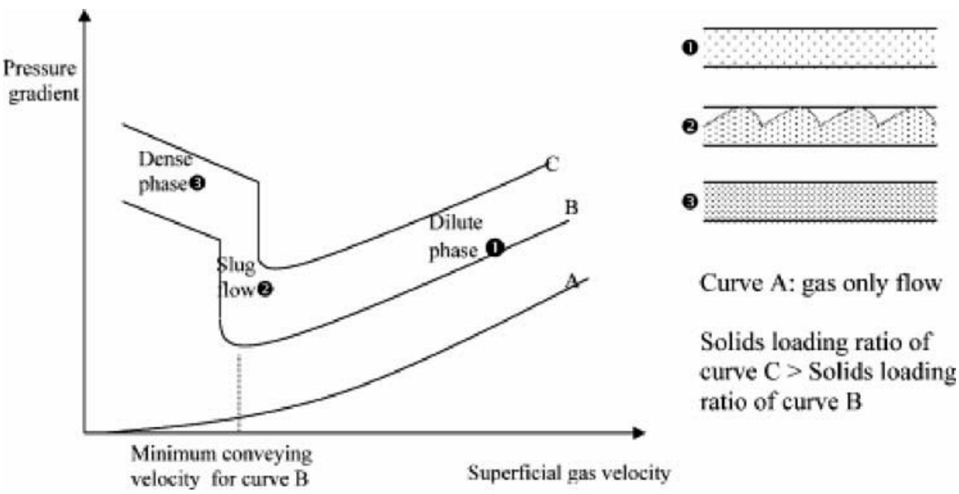
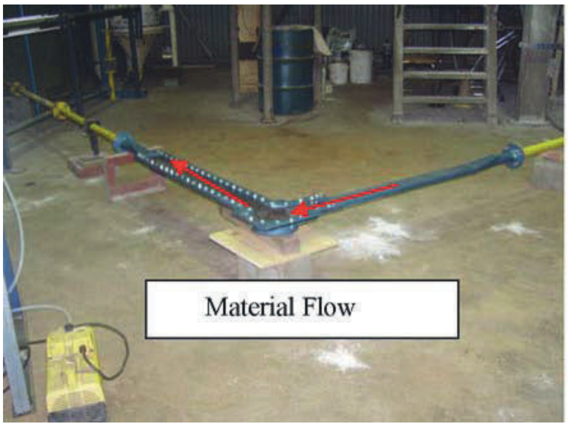


Figure 2: Zenz diagram showing the different phases of pneumatic conveying.

3 VISUALISATION TEST RIG

For a better understanding of the dense phase flow characteristics and its behaviour with changes in conveying parameters, a visualization test rig was designed and constructed at the University of Newcastle. The primary aim of this development was to study the flow patterns in different bends as well as in the particles reacceleration zone after the bends. Another objective of the test rig was to provide a basis for the determination of the critical wear areas with respect to the pipeline configuration and conveying parameters. Although the dense phase conveying was the main focus of the current project, the test rig is designed to study the flow characteristics of both the lean phase and dense phase conveying systems. The following figure (Figure 3) shows the visualization test rig with a circular bend (Botswana bend) installed.



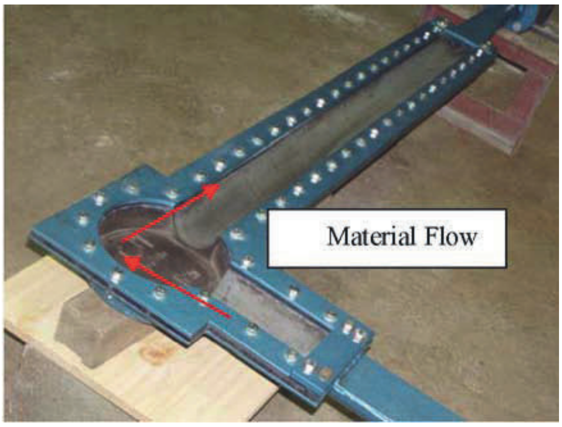


Figure 3: Visualisation test rig with a circular Bend (Botswana Bend) installed.

Alumina with an average particle size of 70 µm was used for the flow analysis in the test rig with varying solids and air mass flow rates used. Changes in the flow patterns inside the bend as well as in the reacceleration zone after the bends were observed and recorded utilising a high speed video at 500 frames per second. The videos were then studied and analysed to investigate the flow patterns and critical wear areas of the flow in the pipeline. In the Figure 4, the flow directions for two different conveying parameters are presented. Flow directions are shown with arrow heads and critical wear areas are shown in solid lines (the right picture has higher gas velocity and no critical wear area).

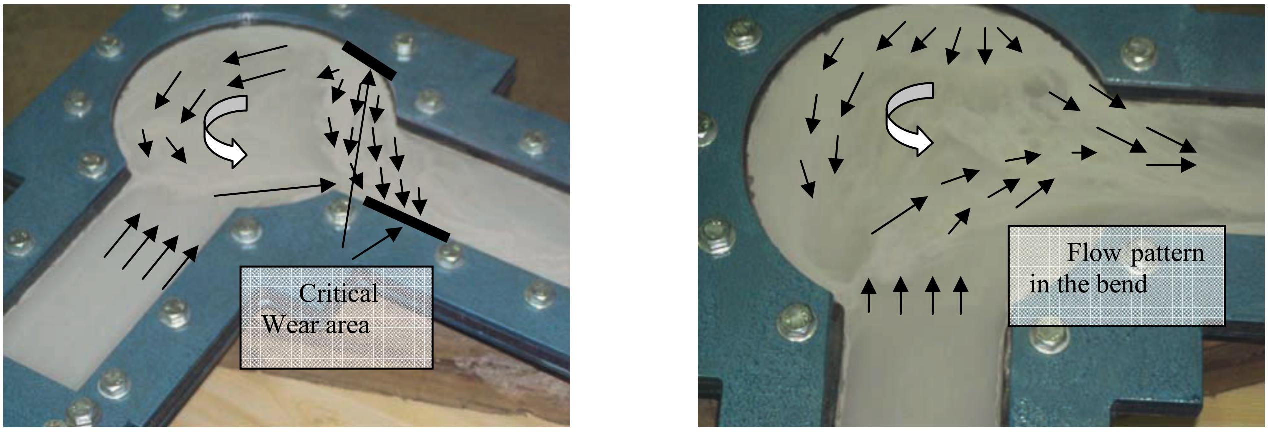


Figure 4: Visualisation test rig and flow patterns in a specific bend.

The two pictures shown in Figure 4 describe the bulk material flow structure for two different sets of conveying parameters. The particle flow in the left picture shows impact on the pipe wall after the bend as well as the particles impact on the bend opposite to the entry. On the other hand, the right picture shows particle streaming through the centre of the pipe with material piled up at the impact point (stagnation point) inside the bend. As the materials impact on the pipeline is reduced in the right picture, the wear in this section is expected to be reduced. This analysis demonstrates the possibility of optimization of the conveying parameters to reduce the wear in the reacceleration zone of the pipeline.

4 FLOW ANALYSIS FOR DETERMINING THE CRITICAL WEAR AREAS IN PNEUMATIC CONVEYING PIPELINE

The visualization test rig clearly showed the structure of the bulk material flow during pneumatic transport through the bends. It was also found that the amount of material that piled up in the bends depended on the solids and air mass flow rates. With higher solid mass flow rate, material piled up in the bend particularly around the impact area which protected the bend from direct impact or erosive wear. On the other hand with increased air mass flow rate, little or no material remained stationary in the bend, which leads to erosive wear due to particle impacts on the bend. After the impact on the bend, the material rebounds and is reaccelerated towards the exit pipe. The flow pattern in the bend clearly shows two different stream paths of the particles; one stream towards the exit and the other in the opposite direction (shown in Figure 4). It is also understood that the reverse flow created in the bend plays an important role in the resultant exit flow from the bend.

From the visualisation analysis, important geometrical features were determined and are modelled as is shown in Figure 5. The geometrical parameters represent the flow patterns in the bend as well as the position of the critical wear areas. The 1st critical wear area occurs directly opposite to the entry pipeline and at higher solid mass flow rates this area is protected by the bulk material build up. On the other hand, if the air velocity is increased, even with the same solid mass flow rate, the material may not build up in this area which will allow erosive wear to occur. The particle stream towards the exit tends to impact on the pipe wall immediately after the bend (2nd critical wear area). The geometrical parameters of distance X and the angles α and β primarily depend on the solid and air mass flow rates in the conveying process. Pressure at the location of the pipeline may also affect the location of the critical wear point due to the variation of the gas velocity.

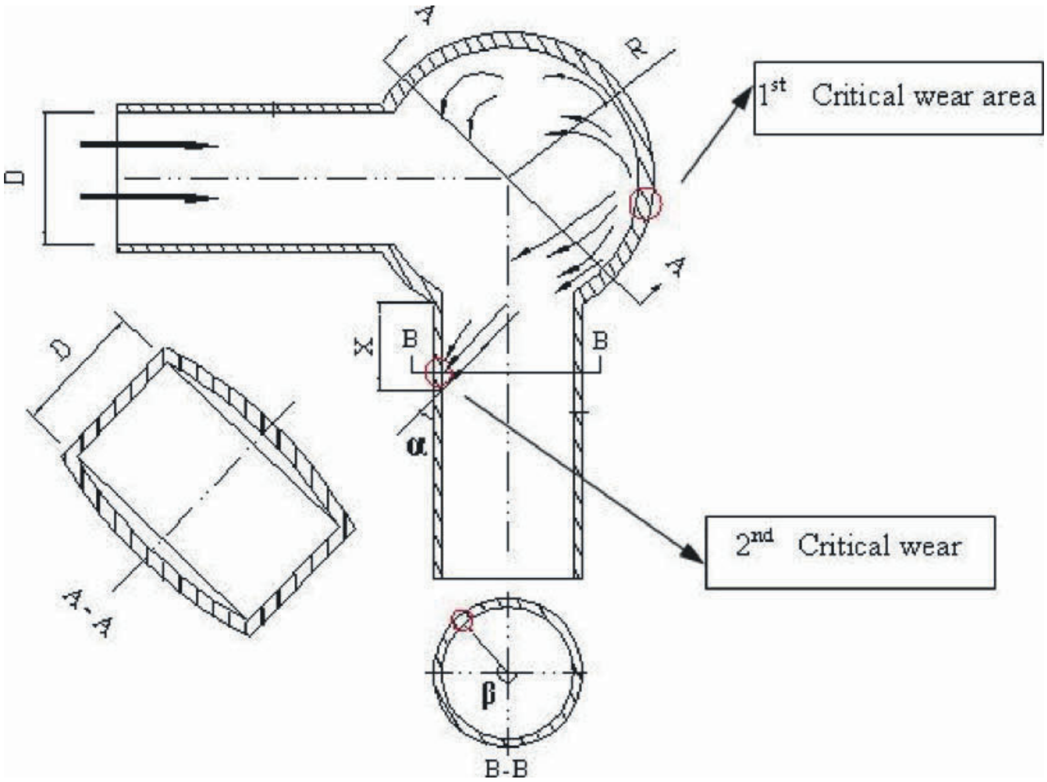


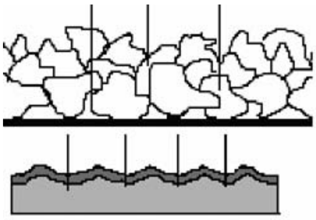
Figure 5: Flow model in the bend based on the high speed video analysis.

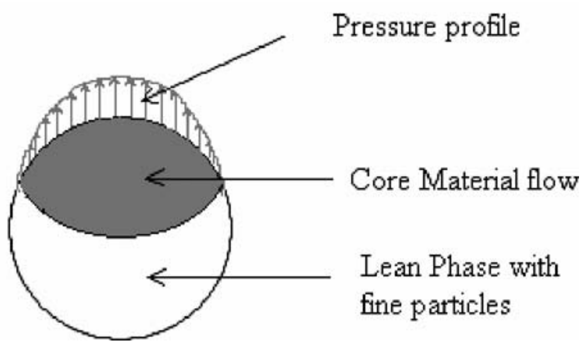
From the analysis of the high speed videos of the flow at different conveying parameters it was clear that the material impact in the 1st critical wear area can be reduced through increased material flow in the bend which increases the material build up in the bend. On the other hand, increased air flow velocity reduced the stagnation material inside the bend which will increase the wear in this critical wear area. Increased air mass flow rate also increases the impact on the 2nd critical wear area which increases the overall wear in the pipeline. With opimisation of the conveying parameters, it is possible to obtain a set of parameters which reduce the impacts on both the pipeline and the bend in order to reduce wear.

5 PARTICLES SEGREGATION

Particle segregation after the bends is a well known phenomenon in pneumatic conveying of particulate materials. Generally a particle jet is formed after the impact in the bend and reaccelerated towards the exit of the bend. In this process, smaller particles mostly follow a stream line with the air flow and larger particles form the particle jet. The particle jet forms a dense core which contains the majority of the material and mostly flows along the pipe wall until being evenly distributed across the pipe cross section further downstream of the bend.

The important aspect of the dense core flow is the associated loading condition that occurs on the pipe wall. The expected surface loading condition (Figure 6) in critical wear area 2 is also an important factor in understanding the severe nature of the wear in this area. This is because as the dense core of bulk material impacts on the pipe surface immediately after the bend, a close packing of the particles generally occurs in the contact area. Although the impacts are within the elastic range, due to the continuous nature of the material flow, the surface does not have enough time to recover from its previous impacts. As a result the surface will accumulate plastic deformation within the individual elastic impacts and cause the surface to work hardened. As the surface is continuously subjected to the impact of the particles as well as sliding of the bulk material, the wear behaviour can be described as impact sliding due to the impact wear superimposed by the sliding. The surface loading and pressure profile in the critical area can be schematically represented as shown in Figure 6.





a) b)

Figure 6. Schematic representation of a) particle loading and surface deformation and b) pressure profile due to bulk material on the internal pipe wall.

The thickness of the work hardened surface layer due to impact sliding depends on the particle size and shape, as well as the pressure fluctuation in the contact area. Work hardening reduces cutting and deformation wear rate but due to the inclusion of alumina into the pipe material from the sliding wear process a transfer film is generated on the surface. The transfer film is characterised by a hardened brittle layer with the effect on the wear characteristics recently discussed detailed in a journal paper [4]. The complex stress cycles in sliding abrasion and impact sliding result in significant alterations of the surface and the subsurface. The changes in material properties in the subsurface layer govern the wear behaviour in the critical wear areas of the pneumatic conveying pipeline. This sub-surface modification of the material was presented by Glardin and Finnie [5] as shown schematically in Figure 7. The fragmented (or transfer) layer (FL) plays a crucial role in the material removal in this region. The transfer layer consists of ultrafine crystallites of the surface material as well as materials from the conveyed bulk material. Once the layer is formed and subjected to rolling fatigue, it is removed through sub surface delamination and cracking.

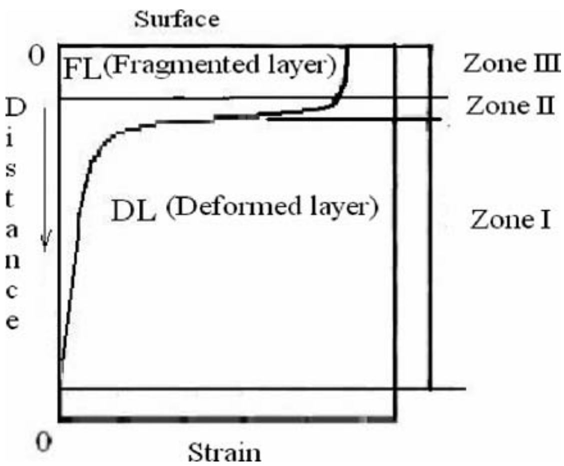


Figure 7. Schematic diagram of the strain distribution below the wear surface and regions defined as 'deformed layer' (DL) and 'fragment layer' (FL).

Particle rolling and sliding has a serious effect on the transfer layer formed on the wear surface. Due to the repeated loading and unloading cycles as a result of the particles rolling, the hardened surface delaminates from the substrate. The particles rolling and sliding conditions can be explained by considering the surface friction and the bulk material internal friction as described in Figure 8 [6]. If the surface friction is higher than the internal friction of the fluidized bulk material, as the case in pneumatic conveying, particle rolling is expected in the critical areas of the pipeline. Fluctuation of pressure and the material flow causes the micro cracks to appear on the delaminated surface layer which is removed through spalling. It is believed that this is the most severe of the material removal mechanisms in the critical wear areas in the pipelines after the bends.

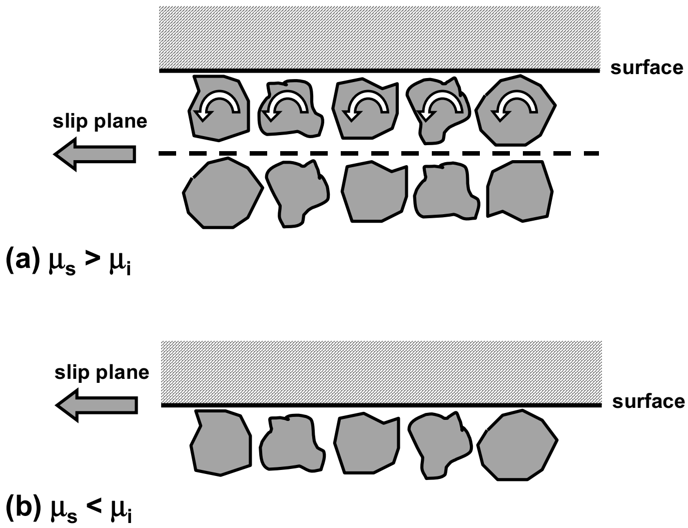


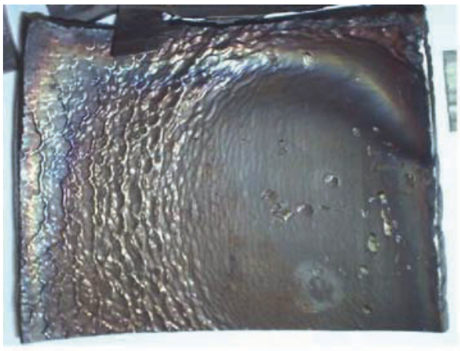
Figure 8. Particle rolling and sliding conditions based on the internal friction (μ_i) and surface friction (μ_s) [6].

6 WEAR MECHANISMS IN INDUSTRIAL PIPELINES

In this section the severe wear areas from an industrial pipeline are discussed and compared with the flow profiles from the visualization test rig. Wear patterns on the samples agreed well with the flow patterns observed during the visualization studies. The locations of the two critical wear areas on the pipeline were also confirmed by the flow analysis in the bends and the straight sections after the bends. Figure 9a and 9b show the bend sections from the dense phase industrial conveying pipeline. The smooth section (figure 9b) represents the area directly opposite to the material entry (stagnation area). The upper right corner of the sample shows a wear groove which represents the area where particles reaccelerate towards the exit of the bend. The wave pattern surrounding the smooth area represent the deformation of surface due to the bulk material movement as seen in the reverse flow during visualization tests. Both the stagnation area and the deformation reverse flow area suffer minimum wear loss from the surface.

Figure 9a shows sections from bends towards the end of the conveying pipeline. In this case, the bend was not protected from the inlet material jet. This is the case when the material does not build up in the bend either due to lower solid mass flow rate or the higher air velocity. In dense phase conveying, although the solid loading ratio remains constant for the whole length of the pipeline, the increase in the air velocity through the pipeline generally means that material does not build up in the bends located towards the end of the pipeline. As a result, the surface opposite to the inlet pipe is subjected to increasing amounts of erosive wear from the particle jet.

a)



b) c)

Figure 9: Critical wear areas as seen on the industrial pipeline and bend samples a) Bend section b) straight pipe section after the bend.

Figure 9c represents the pipe section after the bend showing the critical wear area after the bends. In this case the critical wear area is away from the bend compared to the critical wear area immediately after the bend represented by the patching of the pipe (critical area 2). Figure 9c also shows the spiral nature of the material flow after the bend.

Wear surfaces from the critical wear areas have been studied to investigate the dominant wear mechanisms responsible for material removal in these areas. Detailed study of the wear mechanisms can be found in [5]. One of the very unique characteristics of the wear surfaces observed in this study is the surface cracking and generation of crack network on the wear surface. These features are seen primarily along the bottom of the wear grooves in the severe wear areas in the straight section of pipeline after the bends. Figure 10 shows such a crack network seen on the wear surface.

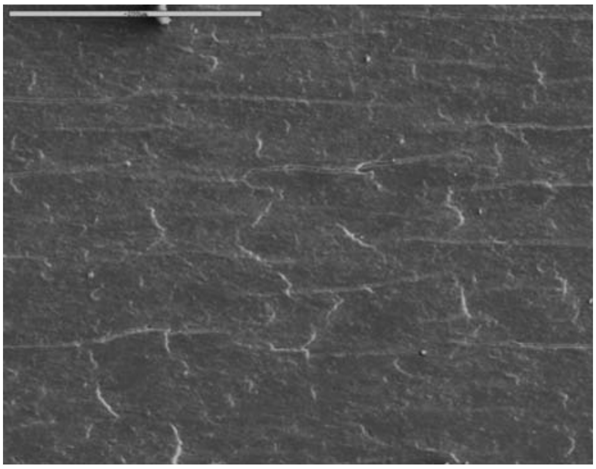


Figure 10. Formation of lateral and longitudinal cracks (material flow from left to right, length bar = 200 μm) [5].

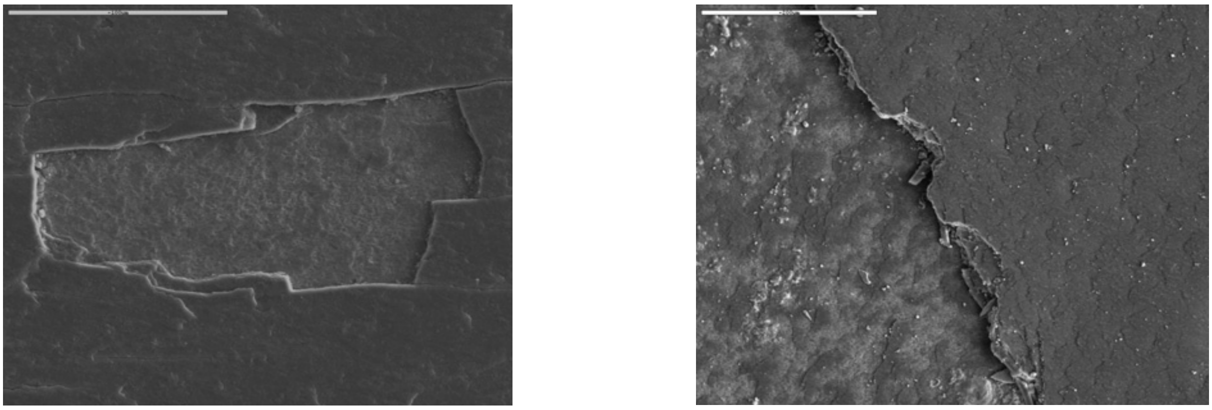
Although surface hardening due to abrasive and erosive wear have been investigated before, these cracking phenomena have not been mentioned in any earlier literature. With careful observation it was found that the cracks are generated preferentially surrounding the rippled structures. One of the possible causes of these cracks can be the surface hardening and the fluctuation of conveying pressure in the pipeline. The cutting and ploughing mechanisms together with the fluctuation of conveying pressure as well as the fluctuation of the rope structured (core flow) material flow causes the surface hardening. The fluctuating pressure may also create a hoop stress condition causing the longitudinal cracks on the surface. As the wall thickness is reduced, any pressure fluctuation in the pipeline can generate a significant amount of hoop stress fluctuation. The effectiveness of the hoop stress increases according to the following formula:

$$\sigma_h = \frac{pr}{t}$$

where σ_h is the hoop stress, p is the pressure inside the pipeline, r is the radius and t in the pipe wall thickness.

It is also possible that a transfer layer is formed on the wear surface which is harder and brittle compared to the substrate material. This transfer layer may also be responsible for the crack formations and material removal through delamination and spalling. Spalling occurs when abrasives flow over the delaminated surface and produce a shear traction force greater than the shear strength between the hardened delaminated layer and the ductile substrate.

Figure 11a shows the delamination characteristics of the surface layer where a cracked surface fragment has been removed. Once one of the surface segments is removed, the surrounding segments are removed relative ease with the new undamaged surface subsequently revealed. Figure 11b shows an area where both the new revealed and the hardened surface is evident from the fracture pattern of the layer. As the material is removed in large fragments, the rate of material removal at this stage is extremely high compared to cutting and deformation mechanisms.



a) b)

Figure 11. a) Formation of cracks around ripples and loss of surface fragment, length scale: 100 μm. b) formation of ripples after removal of the fragmented layer, length scale: 200 μm [4].

The behaviour of a hardened layer on top of a relatively softer substrate had been explained through the characteristics particular to a thin hard coating on a softer substrate. It was shown that if the hard coating is not sufficiently supported by the substrate, it will fail at relatively low contact stress because it can not follow the deformation of the substrate (so called eggshell effect) [7]. Although the hard layer tends to protect the softer substrate from cutting and deformation wear, it fails due to the delamination and subsequent cracking and chipping of the hard layer. Cyclic loading and unloading of the surface causes fracture or fatigue cracks, which ultimately destroys the harder surface layer [8]. Bell *et al* [9] demonstrated that under cyclic loading conditions, fatigue failure is initiated at the coating substrate interface (interfacial fatigue). The mechanism for delaminations and surface cracking was also supported by Jiang et al [10] in the case of hard coating.

From the observation of the industrial pipe wear sections together with the flow patterns from the visualization test rig, it became clear that there is a strong correlation between the conveying parameters and critical wear areas in the pneumatic conveying pipelines. By changing the conveying parameters, different flow patterns have been studied which has enhanced our understanding of the wear situations in the pneumatic conveying pipelines. Based on the visualization studies, and the conceptual model of wear in pneumatic conveying, a cyclic model of wear has been proposed which can predict and material removal rate for optimised process parameters and subsequently predict the remnant life of the pipeline.

7 PROPOSED MODELS FOR WEAR IN PNEUMATIC CONVEYING

With careful consideration of the flow structures in the bends and in the reacceleration zone of the pipeline, the critical wear areas have been defined. Although it is established knowledge that the primary wear mechanisms in pneumatic conveying is erosive (cutting and deformation) wear, work hardening and subsequent delamination and cracking of the surface layer plays a significant role in material removal process. Based on the understanding of the material removal process as supported by the wear mechanism studies from the conveying pipelines, a conceptual model for the wear process has been proposed [6]. The model can be described as follows:

- Step 1: Wear process begins on the ductile surface with impact-sliding: material is removed primarily by cutting and deformation.

- Step 2: Work hardening and transfer film formation of the pipe material takes place according to the particle-wall interaction process: at this stage material removal is restricted to the deformation wear giving rise to the surface hardening effects.

- Step 3: Surface delamination of work hardened surface: due to the surface loading and surface material interaction, subsurface damage accumulation occurs. No material is removed in this stage.

- Step 4: Lateral and longitudinal cracking: as the hardened layer is less supported due to delamination, cracks are initiated around the weak surrounding of the ripples that are formed during the cutting and deformation processes.

- Step 5: Removal of delaminated surface layer exposing the ductile surface underneath: material is removed at a higher rate to remove the layer of harder surface.

- Step 6: Cycle repeats to Step 1.

The reduction in pipeline wall thickness can be determined based on the reduction of wall thickness in three different stages, erosion, sliding and the removal of the hardened layer through crack propagation and spalling. The thickness of material removed during a cycle can be expressed as:

$$t = t_s + t_e + t_c$$

t_s = thickness reduction due to sliding

t_e = thickness reduction due to erosion

t_c = hardened layer thickness due to work hardening

These three events form a cycle in the wear process and *t* (total thickness reduced per cycle) can be associated with a frequency depending on the flow characteristics of the conveying system. Material removal through sliding and erosion can be determined by the established wear models based on mass flow rate, flow velocity and material parameters. The work hardening and transfer film thicknesses can be modelled by material characteristics and the contact process.

8 CONCLUSION

A visulisation test rig was designed for a better understanding of the flow characteristics in pneumatic conveying systems. Analysis of the flow characteristics in the bends as well as the reacceleration zone has been presented with respect in relation to severe wear characteristics of the pipeline. A technique for measurement of particle velocity in the reacceleration zone and impact angles in the critical wear areas has been developed and presented. Finally a technique for a predictive model of wear in the critical region based on a conceptual model for the wear process is also presented.

Possibility of optimisating the conveying process parameters to minimise wear in the critical wear areas has been demonstrated. Although in pneumatic conveying, the solid loading ratio (mass of solids to mass of air) remains constant for the entire length of the pipeline, wear behaviour changes through the pipeline due to the changes in pressure and the particle velocity at different points. As a result, the optimum operating condition at the beginning of the pipeline may be altered towards the end of the pipeline.

Acknowledgement

This work forms part of the research program for the CRC for Integrated Engineering Asset Management.

9 REFERENCES

[1] Ruff, A.W. and Wiederhorn, S.M. Treatise on Mat. Sc. and Tech., in C.M. Pre.ece (ed.) Vol. 16, 1979, p. 69.

[2] Preece, C.M. and Macmillan, N.H. Annu. Rev. mater. Sci. 95 (1997), 12.

[3] Sheldon, G.L. Trans. ASME, J. Basic Engr. 92 (1970), 619-626.

[4] Cenna, AA, Williams, KC, Jones, MG and Page, NW, Wear Mechanisms in Dense Phase Pneumatic Conveying of Alumina, Submitted to the international Journal 'Wear'.

[5] Glardon, R. and Finnie, I., 1981, J. Eng. Mater. Technol., 103, pp. 333-340.

[6] Cenna, AA., Allen, S., Page, NW. and Dastoor, P. 2003, Wear, 254 , pp 581-588.

[7] Meneve, J., Vercammen, K., Dekempeneer, E. and Smeets, J., Surface and Coatings Technology, 94-95, 1997, pp. 476-482.

[8] Holmberg, K., 1992, Surface and Coating Technology, 56, pp. 1-10.

[9] Bell, Mao, K. and Sun, Y., 1998, Surface and Coating Technology, 108-109, pp. 360-368.

[10] Jiang, J., Arnell, R.D. and Tong, J., 1997, Tribology International, Vol. 30, No. 8, pp. 613-625.

ASSET RENEWAL DECISION MODELLING FOR PUBLIC INFRASTRUCTURE ASSETS

SG Pudney [a] Y. Sun [a] L. Ma [a] J. Mathew [a] and M. Mueller [b]

[a] *Cooperative Research Centre for Integrated Engineering Asset Management, School of Engineering Systems, Queensland University of Technology, George Street, Brisbane, Australia*

[b] *MPM Global, Brisbane, Australia.*

Abstract: This paper develops a new approach to stochastic modelling of asset renewal timing decisions using net present value derived from attributable revenue and total cost of ownership. This renewal decision model proposed has the advantage of aligning the efforts of asset custodians to the financial needs of the infrastructure management business. Secondly, we identify the option to perform Front End Engineering (FEE) as part of the asset renewal decision process and use the Black Scholes [1] call option formula to calculate the value of an option to perform FEE. The ratio of option value to option cost is suggested as a means of prioritising front end engineering for renewal projects.

Key Words: Asset renewal, Decision modelling, Real options

1 BACKGROUND

When considering the renewal of public infrastructure assets, we are faced with a unique and challenging scenario. The life of a typical asset is usually in excess of 30 years and sometimes more than 100 years. There is an expectation that the asset custodians will ensure the total cost of ownership of these assets is kept to a minimum whilst simultaneously ensuring agreed service standards are maintained and risks to health and safety and the environment are kept to acceptable levels. Yet this must be achieved against a backdrop of:

- Deterioration of asset reliability and maintainability
- Changing customer service level expectations
- Changing customer demographics and hence fluctuations in demand for service
- Decreasing levels of risk tolerance for public infrastructure owners, operators and their insurers
- Changing community expectations for environmental, financial and social sustainability
- A changing regulatory environment
- Constantly fluctuating levels of inflation and
- Relentless technological change.

It is generally commented that in developed countries aging infrastructure assets are entering the stage where large-scale renewal is imminent. Much of Australia's public infrastructure was built in the post war boom period that lasted from the 1950s through to the late 1970s. Since that time infrastructure development has slowed significantly and it is only in recent times that we are beginning to see an increase in infrastructure capital expenditure through innovative project funding methods. It stands to reason that if most of our infrastructure was built within a short space of time, the need for its renewal will fall due within a concentrated time span.

The peak Australian engineering body, Engineers Australia, has commissioned a series of "report cards" that rate the condition of public infrastructure both at a national level and for each of the states and territories. The summary reports provide a series of subjective grades and comments in a format similar to a school report card whilst the detailed reports discuss aspects of infrastructure condition, performance, governance and management. The American Society of Civil Engineers in the United States and the Institution of Civil Engineers in the UK compile similar reports. In general, these reports contain many references to deteriorating condition and performance of infrastructure assets and concern about the future trends.

It is common to see various forms of public infrastructure reach the point of impending failure before a renewal project commences. In fact, history provides many cases of the remedy being too late. Anecdotal evidence points to a widespread problem faced by infrastructure custodians:

It is difficult to justify a renewal project without the evidence to suggest it is needed but oftentimes by waiting for the evidence to emerge, the time available to plan, prepare and implement the optimal solution is insufficient.

The authors believe quantitative business cases are needed to support the infrastructure renewal projects necessary to build economic growth and competitiveness. This paper defines and demonstrates a new methodology and associated model for systematically building sound business cases for renewal decisions including a method for:

- determining optimal timing of asset renewal based on NPV calculations
- flagging those assets for which renewal is an unsound business proposition

- understanding the value of an option to commence front end engineering (FEE) of a renewal project
- deciding which renewal projects should commence FEE in preparation for possible implementation within the medium term.

2 RENEWAL DECISION MODELLING APPROACHES CURRENTLY AVAILABLE

Typically, the authors of asset renewal decision models have used a Total Cost of Ownership (TCO) calculation at the heart of the decision model. Few have used Net Present Value (NPV) or Internal Rate of Return (IRR) calculations perhaps because of the difficulty of assigning appropriate revenue quantities to assets. Woodhouse [2] suggests that an appropriate approach to asset renewal decision modelling is to maximize life cycle costs and minimise risk simultaneously. Our view is that revenue projections are fundamental to the asset renewal decision and assist asset custodians to understand how assets contribute to or detract from business profit targets.

Renewal decision models that use the prediction of TCO to guide the replacement decision usually include the definition and use of time dependent functions for operating costs, maintenance costs, life extending overhaul costs, replacement costs and salvage costs. Some also propose quantitative consideration of impacts on system users and/or customers and the inclusion of penalties or compensatory claims [3]. Both deterministic and stochastic approaches to decision modelling are seen in the literature. Whilst deterministic models are a useful first step in understanding the influences on a decision to renew, the authors believe that stochastic models are necessary to quantify uncertainty and that this assists the decision maker to justify feasibility studies and other preliminary steps which are necessary in building the case for asset renewal.

The effects of inflation are incorporated into many asset renewal decision models where discounted cash flow calculations are employed as part of a methodology to substantiate decision recommendations. Karsak and Tolga [4] have proposed the use of time dependent differential inflation rates, an approach that recognises some goods and services escalate much more rapidly than others. An appreciation of likely escalation rates and a methodology that enables accurate modelling of cost escalation over the long term certainly assists renewal decisions especially since time spans under consideration are often in the order of decades.

We could not locate any models that specifically incorporated the process view of renewal decisions namely feasibility analysis, front-end engineering, procurement, construction and commissioning as subsections of the asset renewal decision process. We believe that the clear identification of this process and the associated decision points for each stage of the renewal process is essential in building a realistic and flexible working model. For simplicity, we have divided the process into 2 stages – front end engineering and implementation.

We believe that the adequacy of any renewal decision model should be measured by its ability to appropriately deal with the uncertainties of:

- Cost escalation
- Equipment reliability
- Variations in revenue

and at the same time, incorporate flexibility to respond appropriately to these uncertainties at the optimum time. A decision model that considers revenue as well as costs and allows a flexible stepwise approach to renewal decisions is urgently needed.

Traditionally, asset renewal decision models have focussed on two alternatives, either completely replace an existing asset like for like or keep the asset "as is". Roll and Sachish [5] were among the first to consider a third alternative, namely to overhaul the existing asset. A sound asset renewal decision model should investigate both overhaul and replacement strategies against the decision to keep the asset "as is". In this paper, the term "renewal" is meant to encompass both replacement and overhaul possibilities.

3 THE NPV MODEL FOR ASSET RENEWAL DECISIONS

This section describes a proposed asset renewal decision model in detail. We have chosen to maximise the NPV of asset related revenues and costs over a long term period which extends past the design life of the asset. This has the advantage of aligning the decision of asset renewal timing with the financial needs of the infrastructure business.

3.1 General NPV model for use in asset renewal timing decisions

Maximum NPV is the decision objective. The decision variable is time. We have therefore developed a "renewal timing" dependent NPV function that can be used to maximise NPV. This involves calculating NPV for a range of possible renewal dates starting from one year in the future to a date well past the design life of the asset. The NPV for renewal in year R (NPV_R) is calculated as follows:

$$NPV_R = \sum_{i=1}^{R} \mathrm{Re}_i + \sum_{i=R+1}^{N} Rn_i - \sum_{i=1}^{R}(OCe_i + MCe_i) - \sum_{i=R+1}^{N}(OCn_i + MCn_i) - \sum_{i=1}^{R} pe_i CCe_i - \sum_{i=R=1}^{N} pn_i CCn_i - RC \quad (1)$$

where Re is the present value of revenue attributable to the existing asset, Rn is the present value of revenue attributable to the new asset, OCe is the present value of operating costs for the existing asset, MCe is the present value of maintenance costs for the existing asset, OCn is the present value of operating costs for the new asset, MCn is the present value of maintenance costs for the new asset, CCe represents the present value of consequential costs associated with malfunction or failure of the existing asset whilst pe is the associated probability CCn represents the present value of consequential costs associated with malfunction or failure of the existing asset whilst pn is the associated probability and RC is the present value of the renewal capital cost. The value R represents the number of years into the future that renewal takes place (the decision variable) and N represents the number of years into the future that we sum costs. As mentioned previously, this should be well past the original design life as it is often true that an asset is required in service well past its design life.

By calculating NPV_R for each year successively, we can easily determine the maximum NPV and the year R in which this occurs. For cases where no maximum is apparent, revenue may be either decreasing with time (indicating the optimum time for renewal has already passed) or increasing with time (indicating optimal renewal timing is beyond the selected value of N)

3.2 Differential Inflation

Each value in equation (2) is defined in such a way as to take into account differential inflation of goods and services over the forecast review period as proposed originally by Karsak and Tolga [4]. This approach takes into account the fact that different goods and services may escalate at different rates.

Using differential inflation modelling, it is possible to build an accurate forecasting model for each of the time dependent cost components of NPV. In each case it is advisable to separate labour and materials components and use historical data as a guide to predicting future escalation.

3.3 Revenue

The revenue calculation should use an applicable model for calculating "attributable revenue" and differential inflation modelling. "Attributable revenue" is that revenue which can be attributed to a single asset within a system of assets. Whilst it is difficult to propose a general attributable revenue calculation model for all types of public infrastructure, the following aspects should be noted:

It is common for revenue to be broken into two components:

1. An allocation charge – a charge levied for the right to use a certain portion of the service or facilities on offer. As an example, the telephone line rental charge fits into this category.
2. A usage charge – a charge levied for the actual quantity used by the customer. The charges per telephone call or per call duration is an example of a usage charge.

Attributable revenue (as a percentage of total system revenue) can be assigned according to asset usage, throughput, throughput capacity and other asset specific variables.

The attributable revenue model must take account of predicted changes to the customer base both in number and usage patterns with respect to time. The allowable escalation in charges sanctioned by the regulator is an important piece of information that will allow us to predict likely changes in revenue with respect to time. Changes in attributable revenue can also occur when the asset is renewed especially when upgrades or downgrades are being modelled.

We have not included a general model for attributable revenue in this paper since the approach used would necessarily vary between different organisations and business sectors.

3.4 Operating and Maintenance Costs

Maintenance costs should include separate estimates for equipment servicing (such as cleaning, flushing, lubricating, tightening of fasteners and replacement of filter elements), condition monitoring activities (such as the monitoring and analysis of vibration, oil samples, temperatures, thermographic and ultrasonic images, and a range of process variables), preventive maintenance activities or the planned replacement / repair of components, corrective maintenance activities including any action required to return the asset to a state in which it is able to perform its intended function. The use of Weibull functions to model individual failure modes and reliability block diagrams to model asset reliability is recommended. The calculated value for maintenance costs must take into account the different reliability predictions for both the existing asset and the new (or overhauled) asset.

Operating costs should include operating labour, power consumption and a consumables charge for each asset considered. The operating labour is likely to be shared amongst many assets rather than just one; therefore, when modelling operating costs, one should consider the method of distributing labour costs across those assets. The power consumption is directly linked to asset usage. Once again, the calculated value for operating costs must take into account the different power consumption and operating labour predictions for both the existing asset and the new (or overhauled) asset.

3.5 Cost of Consequential Impacts on Stakeholders

Asset failures or malfunctions may have significant impacts on a variety of stakeholders such as employees, customers, the community or the environment. Whilst the probability of such events may be quite small, they may be significant to the decision to renew an asset. We recommend estimating these potential costs and their associated probabilities in all cases and to consider any inherent difference between the old and new (or overhauled) asset.

3.6 Renewal Capital Costs

Estimates for renewal capital costs must include all stages of the project from feasibility and front end engineering through to decommissioning. The capital cost estimate accuracy will depend on the estimating method used. The most accurate method is generally achieved by first compiling a bill of quantities based on a fully detailed design that covers geotechnical, civil, structural, mechanical, electrical, instruments and environmental requirements as indicated by the project scope. The next step is to conduct tender enquiries for the supply of all goods and services required to complete the project. The level of accuracy achieved by following this method is usually better than ± 10%. A less accurate method is to use gross factor analysis and scaling based on similar completed projects for which accurate records are available. The accuracy of this approach is usually in the order of ± 20 to 25%.

Once the estimate has been compiled, the NPV of capital costs for different possible renewal times can be estimated by applying discounted cash flow principles over the appropriate time span. The appropriate discounting rate should be applied to each cost component separately – in line with the previously stated principle of differential inflation.

3.7 Decision criteria other than financial

Decision criteria other than financial may be influential to the decision to renew an asset. Apart from financial, the main decision criteria for infrastructure renewal usually relate to service delivery. If an asset is no longer able to provide service as specified in the Service Level Agreement (SLA) with customers, renewal of the asset must be considered. The types of performance measures used within SLAs include number of service interruptions per annum and duration of service interruptions. Any performance measures with associated targets that are included in SLAs should automatically be used as decision criteria. Whilst these criteria are important to the renewal decision, we do not consider them in detail in this paper but merely recommend that they should be considered alongside the financial decision criteria.

4 STOCHASTIC MODELLING OF THE NPV DECISION AND RENEWAL PROJECT OPTIONS

The deterministic NPV decision model shown in equation (1) is suitable for making estimations about the optimal timing of asset renewals in the short term; however, it does not take into account the volatility of costs and revenues that will occur over time. We recommend stochastic modelling to gain understanding of the magnitude of NPV volatility and real options analysis to determine the value of the front end engineering option.

4.1 Why Stochastic Modelling?

When we have flexibility in the timing and commitment level of our decisions, uncertainty can be a good thing. Stochastic modelling is used to quantify the uncertainty of the renewal project NPV. This uncertainty represents opportunity that is not taken into account with the deterministic NPV calculations.

Flexibility for asset renewal projects comes in the form of options. An option is defined as the right (but not the obligation) to acquire an asset at some future point at a known price. Options in an uncertain environment are valuable. There are methods available for assigning a value to each option. The stepwise nature of the renewal decision process gives the following options:

- Carry out renewal project feasibility study then either: (a) wait for optimal timing to do the rest or (b) abandon the renewal project
- Carry out feasibility study and front end engineering then either: (a) wait for optimal timing to do the rest or (b) abandon the renewal project
- Carry out the entire project now.

4.2 Real Options Analysis

In 1973 Black and Scholes [1] published a seminal paper on a method to price financial options. The Black-Scholes option pricing model was incorporated into a programmable Texas Instruments calculator soon afterwards and it was found that the derivative traders on the floor of the New York stock exchange with access to this pricing methodology via the TI calculator were doing better in their trades than those without.

The model uses six variables to calculate the price of an option using a closed form equation. The variables used in the calculation are:

- The stock price
- The exercise price converting an option to a share
- The dividend yield for the stock
- The volatility of the share price (measured as the standard deviation per annum)
- The time to option expiry
- The risk free interest rate

In subsequent years, a number of management consultants were responsible for applying the Black-Scholes and other financial option pricing models to decisions of many different types. The result was the birth of a new field of endeavour known as Real Options Analysis (ROA). The essence of this new field is summed up in the similarity drawn between financial and real options by Leslie and Michaels [6] shown below:

The six levers of financial and real options	
Financial Option Value Levers	**Real Option Value Levers**
Time to expiry	Time to expiry (T)
Volatility in stock price movements	Volatility in expected cash flows (σ)
Stock price	Present value of expected cash flows (NPV)
Dividends	Value lost over duration of option (nil)
Risk-free interest rate	Risk-free interest rate (rf)
Exercise price	Present value of fixed costs (X)

Figure 1: Analogy between financial and real options

4.3 Calculating the value of an option to carry out FEE

There is often value in doing the FEE of a renewal project early. The upside is that the optimal timing for the renewal project may be earlier than expected in which case the engineering is already done and the renewal can go ahead without delay. The downside risk is that the asset may no longer be needed due to changes in services demand and the engineering for renewal has been done (but this is better than fully completing the renewal and finding out it is redundant).

The value of the option to do FEE is analogous to the value of an American call option for a stock.

The steps in calculating the value of an option to do the FEE are:

1. Estimating the volatility of NPV which is done by:
    a) Collecting relevant data for determining renewal NPV
    b) Calculating revenue and cost components using data collected and then calculate NPV deterministically
    c) Determining the likely range for each variable
    d) Carrying out the sensitivity analysis
    e) For those inputs that contribute significantly to the variation of NPV, determining the distribution type and estimating the likely standard deviation
    f) Determining the likely correlation between input variables
    g) Running a Monte Carlo simulation of renewal project NPV

2 Determining the value of the option to do FEE using the following formula. Refer to figure 1 for an explanation of the variables. Note that Φ is the cumulative standard normal distribution function.

$$Call = NPV_t \Phi(d_1) - Xe^{-rfT}\Phi(d_2) \qquad (2)$$

where $d_1 = \dfrac{\ln(NPV / X) + (rf + \sigma^2 / 2)T}{\sigma\sqrt{T}}$

and $d_2 = d_1 - \sigma\sqrt{T}$

Once the FEE option value has been determined:

3 Compare the cost of FEE with the option value

4 Compare the option value / FEE cost ratio with other projects to help determine priority. Note, additional ways of determining priority may be employed – leading to a combined prioritising system.
5 Decide which projects to carry out FEE using the combined priority.
6 Carry out FEE for those projects that have the highest priority
7 Review options again at a later time and change priorities if necessary

5 APPLICATION TO A WATER DISTRIBUTION SCHEME

Application of the proposed model is briefly presented in this section. Our example is the pipelines for a water distribution scheme installed mainly in the 1960s and 1970s. The scheme consists of approximately 210 kilometers of pipelines which are mainly reinforced concrete (RC). The number of pipeline failures has been increasing in recent years and it has been found that one specific failure mode is dominant, namely joint failure. The asset custodians are currently considering their asset renewal alternatives.

Whilst preventive clamping of joints (an overhaul option) could be used as a means of life extension for the existing RC pipes, this is an expensive option. Technological advancements in pipelines since the 1960s has given rise to a new pipe material, High Density Polyethylene (HDPE) which has several advantages over the original RC pipelines. The cost of manufacture and installation of HDPE pipe is much less than RC pipe and the long-term reliability is likely to be higher than RC. This is the most cost effective renewal alternative but it requires significant engineering input.

The management is concerned at the cost of carrying out pipeline renewals too early; yet leaving renewal too late could result in high maintenance costs and too many service interruptions. There is a significant probability that some of the pipeline assets in this scheme may become redundant due to falling demand. Many farmers in the scheme are selling their land and the crop mix is changing. Some new crops have a much lower requirement for water. It could be a poor decision to renew a pipeline servicing a declining area. A better solution may be to enter into negotiations with landowners about terminating the supply of water and offering some form of compensation.

5.1 Input Data

The following important input data for the decision model was used:

Revenue related data requirements

- Water allocation charge to customers
- Water usage charge to customers
- Water allocation (for asset system)
- Water usage (for asset system)
- Attributable revenue for asset (% of total revenue for system)
- Revenue escalation rate

Operating cost related data requirements

- Operating hours
- Operating labour cost
- Power usage
- Power charge
- Operating consumables cost per annum

Maintenance cost related data requirements

- Failure prediction data of existing asset (for failure mode 1 – wearout and failure mode 2 – random)
    - Weibull shape parameter
    - Characteristic life
    - Failure free period
    - Failure mode 1 as a percentage of all failures (assumed at 100%)
    - Failure mode 2 as a percentage of all failures (assumed at 0%)
- Failure prediction data for new asset (HDPE pipe) – as listed above
- Total number of asset components subjected to potential failure – existing asset
- Total number of asset components subjected to potential failure – new asset
- Cost of corrective maintenance per instance – existing asset
- Cost of corrective maintenance per instance – new asset
- Percentage of labour cost to total cost for corrective maintenance – existing asset
- Percentage of labour cost to total cost for corrective maintenance – new asset

Consequential cost related data requirements

- Probability of consequential costs – existing asset
- Probability of consequential costs – new asset
- Consequential costs per event – existing asset

- Consequential costs per event – new asset

Renewal capital cost related data requirements

- Renewal capital costs
- Decommissioning cost for existing asset
- Percentage of labour cost to total cost for renewal capital cost and decommissioning cost

Financial data

- Consumer price index
- Materials differential inflation rate
- Labour differential inflation rate

Service level data

- Percent of failures resulting in service interruption

Note that all labour costs are inclusive of management and administrative overheads

Some assumptions made were as follows:

- Only one dominant mode of failure (failure mode 1)
- Weibull failure distribution was assumed as the best fit
- Historical failure distribution is representative of future failure distributions
- Failure mode A2 (random failures) considered to be insignificant
- All repairs are "as good as new"
- CPI data from the last 20 years is representative of future data

5.2 Results

The NPV model was used to calculate deterministic NPVs for a range of possible renewal dates each of which were one year apart. The results are shown in figure 2 below. A discussion of the results follows.

The top left hand graph of figure 2 shows zero failures until the end of the expected failure free life after which failures rise steeply. This corresponds to an increase in the peak number of service interruptions per annum (shown at bottom left and an increase in maintenance costs (shown at bottom right). The increase in maintenance costs is the main reason for decreasing NPV (shown at top right).

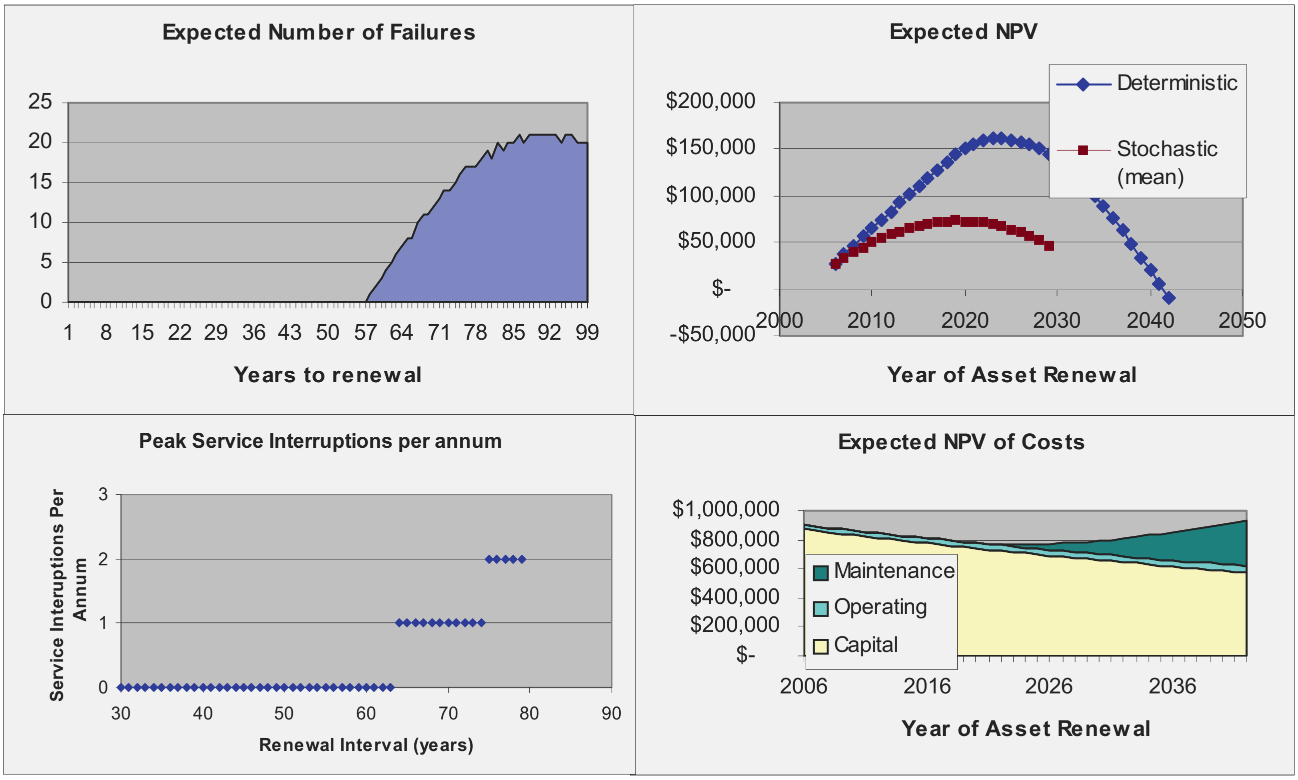


Figure 2: Graphical representation of modelling outputs

Number of service interruptions (SI) is a key measure in the SLA and, as such, is also a key criterion for renewal decisions. If the number of SIs per annum were to rise above 3 for this asset, there is a small probability that the terms of the SLA will be violated. The model predicts that the number of SIs does not rise above 2 and hence there is little likelihood that the SLA will be violated. For completeness, the expected peak number of SIs per annum as a function of the year of renewal is displayed on the bottom left quadrant of figure 2.

NPV peaks when the renewal is scheduled for 2024 (when we examine the deterministic case). When we consider the mean of the Monte Carlo simulation for NPV the optimum renewal date is 2018. This is due to the skewed NPV distribution resulting from the Monte Carlo simulation.

A sensitivity analysis was conducted next which revealed the most the most prolific contributors to variations in both optimal renewal timing and maximum NPV. Not surprisingly, the failure free period is the biggest contributor to variations in optimal renewal timing (figure 3) and it also featured reasonably high on the list of contributors to variations in maximum NPV (figure 4). Eight variables were selected as significant influencers (namely: revenue escalation, consumer price index, water allocation, renewal capital cost, failure distribution failure free period, materials differential inflation rate, labour differential inflation rate and water usage) and appropriate distributions were selected to represent their variation in each case. The correlation between consumer price index and revenue escalation was assumed to be strong positive and was set at 0.8 prior to running a Monte Carlo simulation of 10,000 trials.

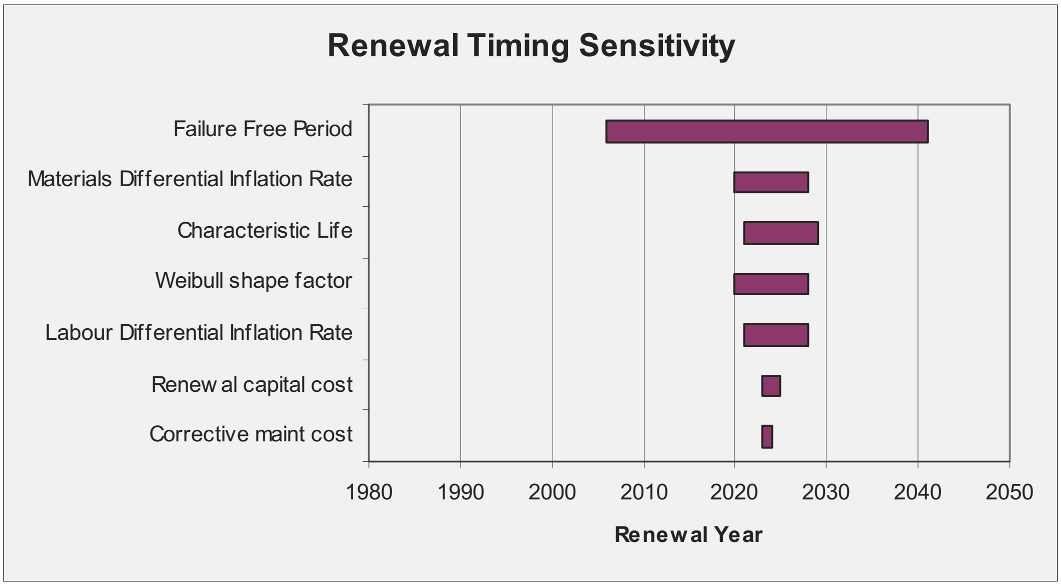


Figure 3: Optimal renewal timing sensitivity analysis results

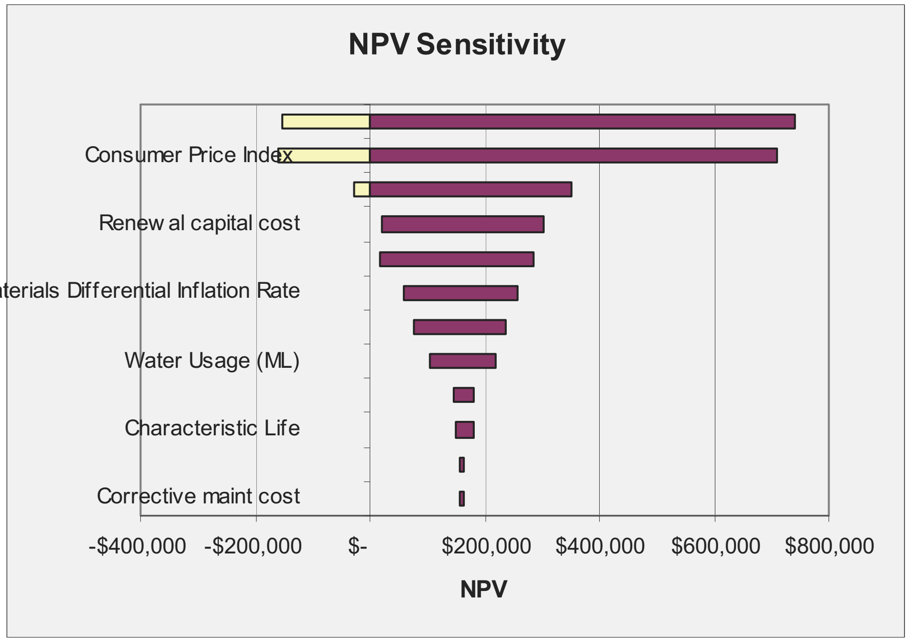


Figure 4: Maximum NPV sensitivity analysis results

The results of the Monte Carlo simulation for renewal in the optimal year (2018) are shown in figure 5. The chart shows that the probability of achieving a positive NPV result by performing the renewal in 2018 is better than 65%. However, the significant possibility of a negative NPV cannot be overlooked. Worse still is the 10% chance that the NPV will be worse than a loss of $205,000. Note that the mean value for the distribution is $74,470 which is significantly different from the deterministic base case figure of $135,629. This is a real example of the way in which a deterministic approach can be misleading.

The results of the Monte Carlo simulation can be used to calculate the NPV volatility. There are several methods. The NPV distribution can be used to calculate volatility, which, in turn, can be used to determine the value of a real option. Using the "Management Assumption" method proposed by Mun [7] the annualised volatility (σ) is given by:

$$\sigma = \frac{SD_{NPV}}{\overline{NPV}} \sqrt{T} \tag{3}$$

Where SD_{NPV} is the standard deviation of the NPV distribution $\overline{NPV}$ is the mean of the NPV distribution and T is the number of periods in one year. In this case T is 1/80 because the data has been summed over 80 years. The resultant NPV volatility is 32.9%.

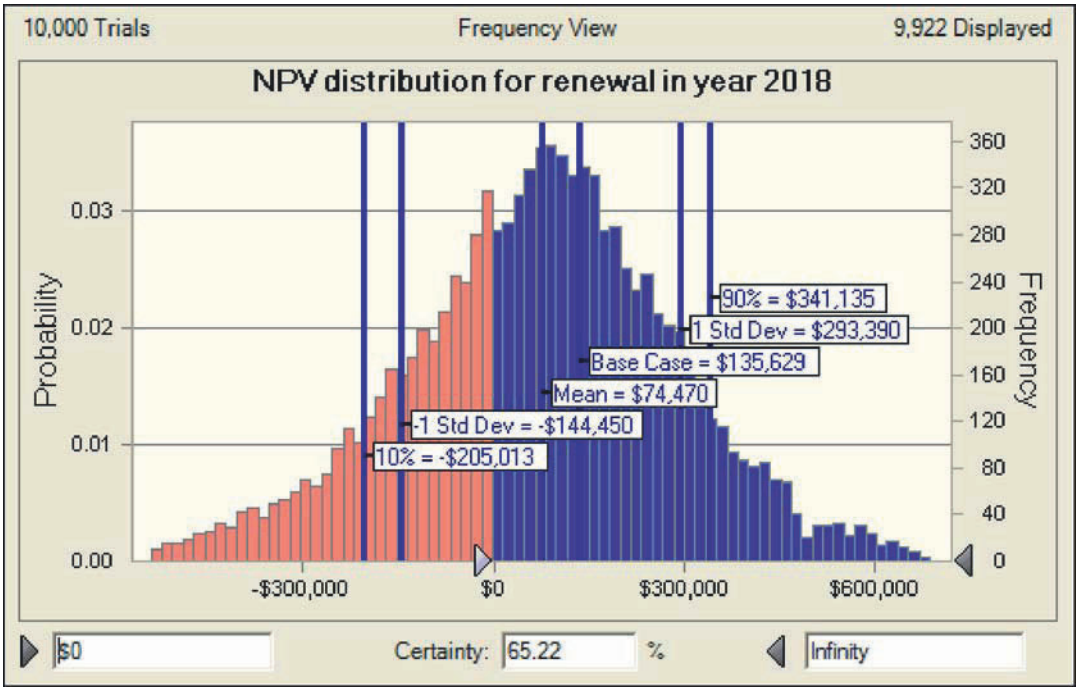


Figure 5: NPV distribution for renewal in year 2018 (optimal)

The Black Scholes equation (2) can now be used to calculate the value of the option to carry out FEE. Assuming that the life of a the engineering design (time to option expiry) is 15 years, the NPV is $74,470, the risk free interest rate is 5% and a present value of fixed costs of $826,453 which is the total renewal project cost less 5% for FEE. A FEE option value of $8,600 is derived. The cost of FEE is estimated at $43,498 hence the ratio of option value to FEE cost is 0.198.

6 ANALYSIS AND DISCUSSION

The NPV of this and other renewal projects can be used to gauge the viability of the project from a business standpoint. In our case, the NPV is likely to be positive at the optimal renewal timing. The mean NPV is a fairly small positive amount of $74,470. There is a reasonable chance that NPV could become negative with lessening demand for water, high inflation, an unexpected increase in capital costs or deteriorating reliability. The regular calculation of NPV for this and other renewal projects would enable the custodian to monitor the need to commence FEE.

Should the NPV predictions become negative, the asset manager should raise concerns with his or her colleagues and discuss the future viability of this asset and alternative approaches to renewal. One alternative could be to negotiate future revenue increases depending on the amount of leeway the regulator allows. Another approach may be to negotiate discontinuing service with the customer in exchange for some form of compensation. In any event, the custodian has a direct measure of financial viability of any proposed renewal projects and this will ultimately lead to sound financial motivation for decisions.

Should the FEE option value become higher than the cost of FEE, this would indicate that carrying out FEE is a sound proposition. There may be further gains in waiting extra time to implement renewal so it would be recommended to watch the NPV closely by monitoring the model inputs and performing regular calculations.

Considering that in our example, the option value is $8,600 as opposed to the cost of FEE being $43,498 (a ratio of 0.198). At this point, there doesn't seem to be much value in performing FEE. It would be better to wait several years when the option value is likely to be much higher. In any case, the FEE option value to cost ratio could be a useful method for prioritising FEE for renewal projects.

7 CONCLUSIONS AND FURTHER RESEARCH

The authors have argued the case for using NPV calculations for supporting asset renewal decisions as opposed to the commonly used Life Cycle Cost or Total Cost of Ownership methodology. The authors strongly believe in using a decision model that helps asset custodians to understand how assets directly contribute to or detract from business profit targets over an extended period. Using a decision model with NPV derived from attributable revenue less costs as the core calculation is one way of achieving this.

The authors have used a case study to illustrate NPV based decision models and the information that can be derived from them. Using the deterministic NPV model the case study examined shows a positive NPV of $135,629 when renewal is scheduled for optimal timing in year 2024. By using the stochastic NPV model, we see that renewal is optimal in the year 2018 and is still cash flow positive ($74,470) but is not as lucrative a proposition as the deterministic calculations would indicate.

The stochastic NPV analysis also indicates a significant probability (almost 35%) of the optimally timed renewal project resulting in a negative NPV with a 10% chance that NPV is worse than $205,000 in the red. The deterministic analysis does not reveal any information about this possible loss scenario. The case study therefore reveals one example where it could be dangerous to rely on deterministic modelling to decide the financial viability and optimal timing of asset renewal.

Noting that significant volatility in revenue and costs is a characteristic of our case study, we introduced the notion that the renewal project may be best tackled in stages. We nominated stage one as FEE (the FEE option) and noted the advantage of performing early FEE was greater surety of information and the ability to act swiftly on changes in project fundamentals. The downside was seen to be limited to the cost of FEE which is only a small proportion of the total cost of renewal. The Black Scholes equation was used to calculate a value for the FEE option ($8,600). This was compared with the estimated FEE cost of $43,498 and our conclusion was that it was not worth spending money to do FEE at this stage.

The authors therefore conclude more generally that:

- NPV renewal decision modelling is a useful approach in that it focuses asset custodians to serving the financial needs of the infrastructure management business.
- Our case study shows that whilst the deterministic NPV model can be used to obtain approximate timing for optimal renewal, it provides no information about downside risk and is therefore recommended as a guide only.
- Stochastic NPV modelling can be used to reveal the risk of a loss scenario. In our case 35% probability of loss if the asset is renewed at the optimal timing.
- The FEE option to cost ratio is one measure that could indicate those renewal projects for which FEE is justified. A ratio above 1 indicates good value for money. Note that there are other ways of justifying renewal projects, for example on the basis of likely violation of service level agreement.

The authors believe that there is scope for applying stochastic NPV and ROA methods to asset management decisions other than renewal timing.

8 REFERENCES

[1] Black F and S. M, *The pricing of options and corporate liabilities.* Journal of Political Economy, 1973. **81**: p. 637 - 659.

[2] Woodhouse J, *Asset management decision making*. 2001, The Woodhouse Partnership Ltd.

[3] Dwight R, et al. *Decision models for capital replacement*. in *The International Conference of Maintenance Societies*. 2005. Hobart Australia: The Maintenance Engineering Society of Australia.

[4] Karsak EE and T. E, *An overhaul-replacement model for equipment subject to technological change in an inflation prone economy.* International Journal of Production Economics, 1998. **56**: p. 291 - 301.

[5] Roll Y and S. A, *Combined overhaul and replacement policies for deteriorating equipment.* Journal of Operations Research Society of Japan, 1978. **21**(2): p. 274 - 286.

[6] Leslie KJ and M. MP, *The real power of real options.* The McKinsey Quarterly, 2000. **3**: p. 4 - 23.

[7] Mun J, *Real Options Analysis*. 2002, New Jersey: John Wiley & Sons.

COMMON PITFALLS WITH SAP™-BASED PLANT MAINTENANCE SYSTEMS

Leonie Gouws & Johan Gouws

Melikon Pty Ltd (www.melikon.com), Perth, Western Australia

Abstract: This paper discusses some of the most common pitfalls which had been encountered by the authors, over more than ten years, with SAP-based plant maintenance systems. The experience is drawn from maintenance engineering consulting for a variety of companies where the SAP Plant Maintenance module is used as the integrated computerised maintenance management system. The companies span the defence-, oil and gas-, mining-, and manufacturing industries in Australia and South Africa. The aims of this paper are to identify five of the most commonly encountered problem areas with the SAP Plant Maintenance system; discuss the main effects of these problems on the management of the maintenance process; propose some remedies to improve the effectiveness and efficiency of maintenance relying on the SAP PM module; and identify potential problems to look out for when implementing the proposed remedies. Although this paper is not the *silver bullet* to slay all SAP PM related problems, it will assist maintenance managers and reliability engineers to better understand how and why things can go wrong with their SAP PM systems, and where to look for solutions to alleviate these problems.

Keywords: SAP PM, Plant Maintenance, Computerised Maintenance Management System (CMMS)

1 INTRODUCTION

As part of consulting work in maintenance management optimisation, the authors had been involved in the implementation, use, review and upgrading of the SAP PM module for various companies in the defence-, oil and gas-, mining-, and manufacturing industries in Australia and South Africa, for more than 10 years. During this time, the following were identified as five of the most common areas in which problems develop which hamper effective and efficient use of the SAP PM system:

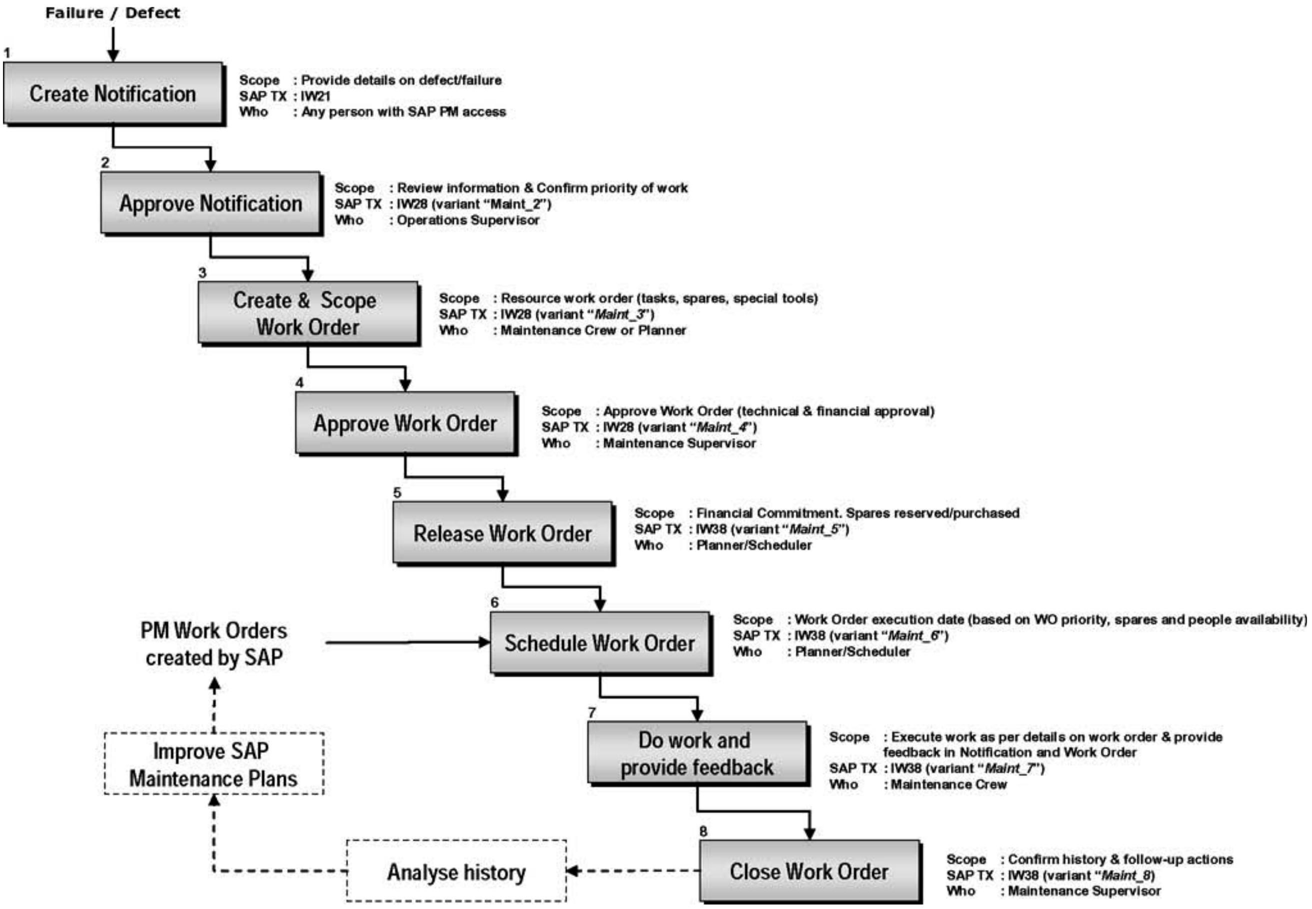
- Maintenance Work Flow Processes;
- Work Centre Structures;
- Work Prioritisation;
- Damage and Cause Codes;
- Master Data Maintenance.

The main issues to look out for in each of these areas are discussed in this paper, as well as some remedies to improve maintenance management based on the SAP PM module. Potential problems to look out for when implementing the proposed remedies are also pointed out. This paper is not intended as a *cure-all* or a *silver bullet* to slay all SAP PM related problems, but it is intended to assist maintenance managers and reliability engineers to better understand how and why things can go wrong with their SAP PM systems, and where to start looking for solutions to alleviate these problems.

2 SAP PM PROBLEM AREAS

2.1 Maintenance Work Flow Process

The Maintenance Work Flow Process (ref [1]) is normally presented as an activity-based mapping, using block diagrams to define the activities involved from identifying a requirement for maintenance to execution of the maintenance activity and close-out of the job. For the Maintenance Work Flow Process to be utilised as an effective communication tool, it should contain sufficient information to define the activities in terms of *what is required*, *who is responsible for executing the activity*, *when the activity is required*, and *how to execute the activity* in the computerised maintenance management system (CMMS). When initially developing the Maintenance Work Flow Process, it should be determined what value the activity will add, how long it should take to execute, how it will interface with other processes, and what risks will likely be associated with the activity. Figure 1 illustrates a typical Maintenance Work Flow Process containing this basic information.



(Source: Author)

The following are common mistakes made during the composition of Maintenance Work Flow Processes:

- Maintenance Work Flow Processes too complex. Many companies use the business processes developed during the implementation of SAP PM, as training tools, but these processes are normally too detailed and complex for training purposes. One case was encountered where a company used a work flow process containing 75 activities, to explain responsibilities to their new SAP PM users. With such a complex business process, the users become confused and frustrated, and invariably they then consider SAP PM as "*user-unfriendly*" and "*difficult to use*". Once such negativity has taken hold, it is very difficult to overcome. Therefore, it is important that maintenance work flow processes should be kept simple. Processes containing more than ten activities are often not well received by end-users.

- Maintenance Work Flow Processes not used during pre go-live training. Sometimes end-user training material is developed separately from the Maintenance Work Flow Processes. This can introduce a lot of confusion and resistance against the new SAP PM module, because there is no clear link between the training material and the Maintenance Work Flow Process. After go-live, users in an attempt to link their SAP PM training material with the work flow end up paging from one training exercise to another, adapting rules on the run. Also, without the Maintenance Work Flow Processes which indicate individual responsibilities, end-users find it difficult to identify what sections of the training are applicable to them. The most efficient training material is mapped to, and details individual steps in the Maintenance Work Flow Process. There is thus a clear link between the training exercises ("how") and the Maintenance Work Flow Processes ("what and who"). The relevant training exercise should also be indicated clearly on the work flow process. Many companies use the term "Quick Reference Guides" (QRGs) and use that in the numbering of the training exercises (e.g. QRG_Maint_1 "Create Notification"). By frequently referring to the relevant QRGs during training, users get familiar with how and where to find the relevant instructions again in their training manuals, later on.

- Insufficient SAP variants to support Maintenance Work Flow Processes. End-users are often expected to develop their own SAP variants (online operational reports using SAP's list editing capabilities) to identify notifications and work orders in the different parts of the work flow process. However, this practise is not advisable for the following reasons: (i) Most new SAP users find SAP's online report development feature difficult to use. (ii) New users do not have the detailed understanding of how the notification and work order system and user status codes change throughout the work flow process; and this knowledge is essential in the development of variants that will contain the correct work orders and notification in the listings. (iii) It is essential for all users to have the same layout of the report. This includes the sorting sequence of the notifications and work orders from highest importance to lowest importance, and the columns provided in the report. (iv) Any changes to the work flow process after go-live, such as the introduction of a new approval step, can be managed with ease if there are only a few generic reports to update. The users will not even be aware of the changes, which will enhance their perception that the system or processes are stable. (v) There should be a clear link between the Maintenance Work Flow Process, the QRGs and the variants, by selecting appropriate standardised names for the variants and ORGs.

It is thus unwise - and might even be a dangerous practise - to require users to develop their own reports in SAP. It is considered essential for the safe and effective management of the maintenance work to provide the users with generic SAP variants and generic layouts for each activity in the work flow process - and to provide the users with the opportunity to run these variants during training.

Insufficient decision gates. There should be a number of technical and financial decision gates in a Maintenance Work Flow Process. The technical gates provide the opportunity to ensure that the work order or notification contains the correct information; while financial gates provide the opportunity to ensure that there are sufficient funds available to do the specified maintenance. The following are gates that can be designed into the Maintenance Work Flow Process:

- Approving the notification – where a supervisor agrees that the maintenance is required, that the priority of the job is correct, that the correct functional location has been selected, and that the notification contains sufficient details to ensure that the notification can be processed further without continuous requests for information from the creator of the notification.
- Approving the work order – where a supervisor reviews the fully scoped work order to ensure that the work order contains the correct labour, spares, special tools, documentation, etc. Apart from ensuring that the scope is acceptable, there is also the opportunity to approve the financial commitment.
- Releasing the work order – this activity initiates the process of obtaining the resources as specified on the work order (i.e. spares, labour, special tools, etc.) on the right date as specified on the work order. The planner/scheduler is normally the person releasing the work order as he/she can then control when the spares are reserved or the purchasing process will start.
- Technical closure of the work order – which implies that the job had been executed as specified in the work order, that all the history had been completed and that all follow-up work had been identified. This is normally done as a final check by the maintenance supervisor.

In many organisations, the maintenance crew are the only people involved from the creation of the notification (activity 1 in Figure 1) to closure of the work order (activity 8 in Figure 1). This can be an expensive and potentially dangerous practise, because in these cases, the operations- and the maintenance supervisors are not actively involved in the maintenance work flow process to confirm the priority of the jobs or the scope of the work. Delegating the actual maintenance tasks to the maintenance crew does not mean delegating the responsibility for decision-making too. To ensure effective decision-making regarding the work flow process, it is important to limit the number of decision gates; and to clearly allocate the responsibility for each decision gate to an authorised, well-trained decision maker.

- Vague guidelines on capturing maintenance history. Activity 7 of the work flow process in Figure 1 includes capturing of the maintenance history as part of the close-out process. The following information is normally required:
  - In coded format, “what was wrong”, “what caused it”, “to what section of the object was the damage limited” and “what was done to address the damage” (i.e. damage, cause codes, object codes, and activity codes on the notification).
  - Brief description of what was done, and what follow-up work is required (normally in the *notification long text*).
  - The time spent to execute the job (part of the time confirmation).

The guidelines on where exactly to capture the maintenance history is not always clear to the SAP users. Long text in SAP’s PM module can, for example, be captured in five independent long text fields in SAP: (i) the notification long text, (ii) the work order long text, (iii) the confirmation long text, (iv) the cause code long text, and (v) the operations long text. Without proper guidelines, the maintenance crew tend to enter detailed descriptions of the work undertaken in the confirmation long text – a field often not even reviewed by maintenance data analysts.

Another problem often encountered relates to the time confirmations: Should only active tool time be confirmed, or should the time to obtain permits, spares and relevant documentation also be included? Should the time to resource the work order also be confirmed? Is there a limit to the deviations allowed between what was planned on the work order, and what was finally confirmed?

Data analysts need to understand the relevant issues and uncertainties regarding the capturing of maintenance history before attempting to analyse the data. Without compensating for these issues, conclusions will be inaccurate and corrective actions ineffective.

- Effectiveness of Maintenance Work Flow Process not monitored. In the majority of companies, there are often periods with more maintenance work than people to do the work. In such cases a backlog of work orders will result, which needs to be understood and actively managed. There are certain activities in the Maintenance Work Flow Process where a backlog is not advisable – e.g. approval of notifications (activity 2), creation of the work order (activity 3), and technical closure of work orders (activity 8). Backlog should not be allowed for activities 2 and 3, because of the risk that high priority work orders might not be executed in time. Backlog in preventive work (PM) orders in activity 8 might stop the generation of future PM work orders (depending on how the maintenance plans were set up in SAP). To ensure that backlog is understood and properly managed, online operational SAP reports for each of the activities in the Maintenance Work Flow Process must be executed on a regular basis by the maintenance supervisor or planner. Many companies fail to do this, and often only after a major failure, do they realise the value of running these reports regularly.
- Poor understanding between the Plant Maintenance (PM) Work Flow and Materials Management (MM) Work Flow. When asking maintenance crew and planners to list their biggest frustrations in terms of the Maintenance Work Flow, difficulty to find out when spare parts will be onsite often feature high on the list. Part of the problem lies in inaccurate cataloguing of the spares in SAP, but the biggest problem is a lack of knowledge of the materials management (MM) module in SAP. The following questions are often encountered:
  - How do I find out whether a material (spare) is catalogued (i.e. whether the details of the material is in SAP)?
  - Is the material attached to a Bill-of-Material somewhere?
  - How do I known how many items are in stock, and where they are?
  - What is the expected delivery time for the non-stocked item (i.e. items purchased using purchase requisitions and purchase orders)?
  - How do I know where the material is in the process (e.g. has the purchase request been turned into a “Request for Quote”, has the purchase order been created, etc.?)

  These questions are difficult to address during initial SAP PM training, since it contributes to information-overload. This is one of the reasons why follow-up training is essential; and why follow-up training for SAP PM users should also introduce them to the SAP MM module.

2.2 Work Centre Structures

In SAP Plant Maintenance, there are two different types of work centres [2]: *Main Work Centres* and *Work Centres*.

Main Work Centres generally define (on the header of the work order) the person/department responsible for ensuring that the specified maintenance task is carried out. *Work Centres* are used to specify the crafts required to perform the specified tasks. Work centres are an important means of integrating with other modules in SAP such as the Financial and Controlling modules (FI/CO), Production and Planning module (PP) and Human Resources module (HR). Refer to Figure 2 for an illustration of a typical work centre structure.

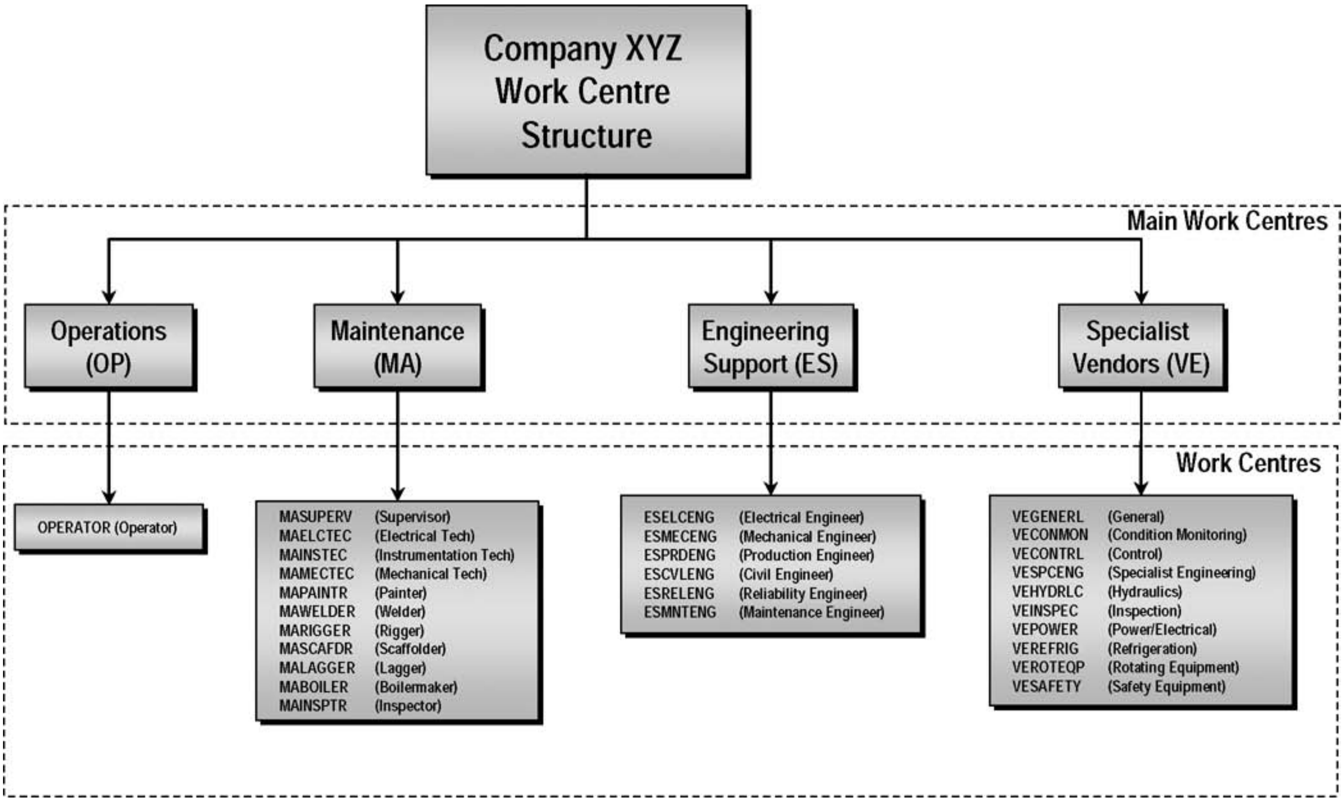


Figure 2. Typical Work Centre Structure

(Source: Author)

A major mistake made by companies is to underestimate the importance of the work centre structure during the initial configuration of SAP. Companies tend to put too little thought into the work centre structure, not realising that the highly integrated nature of work centres makes it extremely difficult (though not impossible) and risky to update the work centre structure once SAP is operational. This is especially true where there is a requirement to combine work centres at a later stage. Another common mistake is to have one person working under two or more different work centres. This complicates capacity planning, while it adds very little additional benefit. The third most common mistake is to not use the main work centre prefix (e.g. MA for *Maintenance*) in subordinate work centres (e.g. *MAMECTEC* for *Mechanical Technician*). This complicates the development of variants used to manage work orders throughout the Maintenance Work Flow Process, and in general the searching and sorting of work orders by work centre.

2.3 Work Prioritisation

Levitt **Error! Reference source not found.** has the following to say on work order priority: "Priority helps assign work where there is more work than people. It ensures that vital work is not overlooked in the rush […]. Priority systems have the habit of being abused so that users can get their work done faster." Each company must decide on the criteria to be used when determining the priority of a job, and the criteria should preferably align with the corporate risk matrix. Figure 3 illustrates a typical work order priority matrix taking into account the actual or potential effect of the defect on production, safety, environment and susceptibility to degradation. Other factors that might also be included in a priority matrix are product quality, historical failure rate, and the age of the work order. The company should agree on acceptable response times for particular work order priorities (e.g. "Complete Work Order within 14 days" for priority 2 work orders). These response times give the planner a window of opportunity in which the work order can be scheduled.

The relative ranking of work orders in each step of the Maintenance Work Flow Process will be based on the priority of the work orders, making it essential to allocate the priority correctly during the approval of the notification. The operations supervisor is normally the best person to determine the effect of a defect on safety, production, etc.; and this person should have a sound understanding of the priority matrix in order to make sensible decisions regarding work priorities.

Probability of failure

Low → High

Impact of failure	Safety	Environment	Production	Condition will not worsen over time	Condition might worsen over time	Condition will worsen over time
High	Serious injuries / fatalities	Major impact	Plant shutdown	3	2	1
	Minor injuries	Minor impact	Process shutdown	3	3	2
	Slight injuries	Slight impact	Equipment shutdown	4	3	3
Low	No injuries	No impact	No impact	4	4	3

Agreed reaction time based on Work Order Priority:

Priority 1 → Will complete WO within 24 hours
Priority 2 → Will complete WO within 14 days
Priority 3 → Will complete WO within 2 Month
Priority 4 → Will complete WO within 6 Months

Figure 3. Typical Work Prioritisation Matrix

(Source [5]: Fundamentals of Software Project Engineering)

The most common problems experienced in terms of the work prioritisation are:

- Priority based on required reaction time instead of impact on plant. In many organisations, the priority of the job is set by deciding when the job should be executed. The required schedule, instead of the consequence of the defect on the plant (as determined by criteria such production, safety and environment) thus wrongly becomes the driver of the priority. Although people instinctively include criteria such as safety in their reasoning when determining the priority of the job, this is done indirectly and will differ vastly from one user to another. As backlog increases, people also tend to increase the work order priority in an attempt to get their jobs done quicker. Therefore it is often found that the majority of the work ends up as priority 2, making it impossible to identify the really critical jobs amongst the "clutter".

- The priority of the work orders are allocated by maintenance crew or planners. Maintenance crew and planners often do not have the right exposure, experience, and "view of the bigger picture" of their plant to accurately allocate the priority of the work order. They tend to base the priority on when they think the work can or should be completed. This is based on how available the spares are, how available the resources are, and how long the job will take to complete. Smaller jobs often get the higher the priority, since these are easier to get "out of the way". The person that will be the most affected by the incorrect allocation of the priority is the operations supervisor, being ultimately responsible for maintaining production. It is therefore advisable that operations supervisor, supported by the maintenance crew, thus decide on the priority of the work order.

2.4 Damage and Cause Codes

Maintenance history is captured by selecting appropriate damage and cause codes from a list of codes provided by the SAP system. The SAP PM system can be configured to provide specific codes for specific types of functional locations / objects. The maintenance history is normally captured by the maintenance technician as part of the close-out of the job. The following two pitfalls have been the most commonly encountered, regarding the configuration of the damage and cause codes:

- List of codes to select from too elaborate. Many companies opt for elaborate and very detailed lists of damage and cause codes - sometimes even multi-tiered lists. One case was encountered where groups of codes contain an average of ten first-tier codes, with six to ten second-tier codes for each of the first-tier codes. The people who have to capture the maintenance history must therefore select the appropriate code from sometimes more that one hundred options! The intent of the detailed coding structure is to cater for all possible options in an attempt to gather comprehensive and accurate history. However, the result is the exact opposite. Maintenance technicians normally enter these codes at the end of a shift, when they are tired and hungry. When presented with a choice of up to one hundred codes for each job, they will almost always limit their selection to the same three codes – irrespective of the actual damage and cause.

When analysing the maintenance history provided of long text in work orders and notification, the data analyst must always get an understanding of the choices provided and the conditions under which this data is normally entered. The fact that details have been entered does not automatically imply that the data entered is accurate. By getting an understanding of the source of the data, the analyst might prevent incorrect conclusions based on detailed analyses using invalid data.

Levitt [3] suggest that preventive maintenance task lists should aim at preventing the *most likely*, *most expensive* and *most dangerous* types of failures. This principle can also be applied with great success to selecting the damage and cause codes to be provided in SAP for each type of object. Damage and cause codes should thus be focussed on identifying the most likely, most expensive and most dangerous failures. A good guideline to aim for when deciding how many codes should be provided per object type comes from Miller [4], who describes the capacity of short-term human memory at seven items plus or minus two, depending on the individual. Further study revealed that when presenting bulleted information in groups of the "magical number seven" the reader "sees" the information as a picture instead of words requiring reading. Limiting damage and cause codes to seven or less therefore makes it easier for the maintenance technicians to assimilate the information and to select the most appropriate code – resulting in more accurate maintenance history being captured.

- Providing generic codes that are not specific to the type of functional location/object. Some companies use generic high-level damage and cause codes instead of developing object-specific codes. Generic codes would, for example, be provided only for groups of objects classified as "mechanical components" or "electrical components". The problem with these generic codes is two-fold: (i) the list of codes to select from is generally very long to allow for all possible options (similar problem as discussed above); and (ii) the codes are not applicable to specific problems encountered, making it difficult to select "the most appropriate code".

2.5 Master Data Maintenance

Plant Maintenance "Master data" can be defined as data not changing frequently during the day-to-day operation of the plant, and form the basic building blocks of the SAP PM system. Examples of PM master data are functional locations, equipment, bill of material (BOM), work centres, measuring points, maintenance task lists, and maintenance plans. Changes to PM master data should go through a rigorous process of review and approval by the relevant custodians of the data before implementing the changes. Refer to Figure 4 for a typical Master Data Management Process.

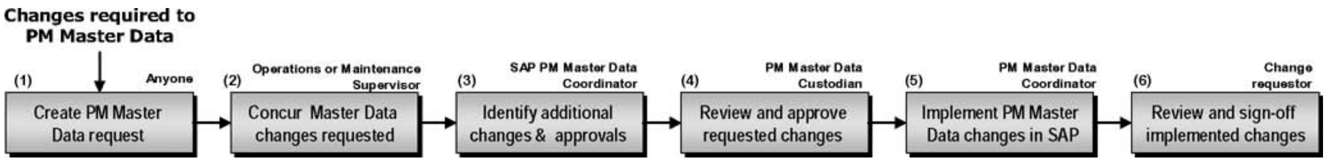


Figure 4. Typical PM Master Data Management Process

(Source: Author)

There is a need for a *SAP PM Master Data Coordinator* to facilitate the whole data update process. The SAP Data Coordinator has to be skilled in maintenance and SAP PM, and has the responsibility of ensuring that the scoping of the change is comprehensive and implementation of changes accurate. A user might for example have specified an increase in the regular inspection interval for a particular functional location by only quoting the relevant PM work order number and the required inspection interval. It is then up to the SAP Data Coordinator to determine the impact of this change on other maintenance plans, whether there are job procedures to update, whether the scheduling parameters on the maintenance plan are still applicable (e.g. is the period between creation of PM work order and required execution date of the work order still valid), etc. The SAP Data Coordinator must also identify the relevant custodian to review and approve the requested changes.

Unfortunately the importance of the correct management and updating of SAP PM master data is often underestimated. This leads to numerous problems, including:

- No SAP PM Master Data Coordinator in the Master Data Change Process. Some companies do not use a SAP PM Master Data Coordinator, but allocate this responsibility to a Master Data Administrator role (i.e. someone with SAP PM skills but with limited or no knowledge of maintenance) who might not understand the intent and full scope of the requested changes. The end-result is master data with numerous errors and inconsistencies.

- Insufficient review of requested changes by the Master Data Custodian. It is expected of the SAP Master Data Custodian (normally an engineer per discipline) to review and approve the requested changes. Unfortunately this person normally also does not have sufficient knowledge of the SAP PM system outside notification and work order management – and does not know how to interrogate SAP in order to gather various bits of information. Changes are then approved without proper review. The solution is regular follow-up training sessions or demonstrations – no longer than two hours per session as the engineers find it difficult to allocate big chunks of their day (and sometimes also big chunks of motivation!) to SAP training sessions.

3 CONCLUSION

In this article five of the most common problem areas in the use of the SAP Plant Maintenance module were identified and discussed. The list is by no means comprehensive, but provides valuable insight for other practitioners in this field. Although some of the problem areas discussed above might be easy to fix (e.g. development of a priority matrix), the difficulty does not only lie in development of the solution, but also in finding a custodian prepared to promote these changes. Implementing changes to the SAP PM system might require users to be retrained, supporting documentation to be updated, users' SAP access to be changed, additional reports to be developed, cause codes to be streamlined, data management to be upgraded, etc. In order to ensure effective and efficient plant maintenance, it is essential that SAP PM, as one of the tools used in this process, is configured and used as well as possible. This requires continuous attention to how SAP PM can be better utilised, instead of an "implement and forget strategy".

4 REFERENCES

[1] Campbell JD. (1995) Uptime: Strategies for Excellence in Maintenance Management. Portland, Oregon: Productivity Press.

[2] Stengl B. & Ematinger R. (2000) SAP R/3 Plant Maintenance – Making it work for your business. SAP™ Press, Addison Wesley

[3] Levitt J. (2003) Complete Guide to Preventive and Predictive Maintenance. New York, NY: Industrial Press Inc.

[4] Miller GA. (1956) The Magical Number Seven, Plus or Minus Two: Some limits on Our Capacity for Processing Information. *The Psychological Review*, vol. 63, pp.81-97 (available at http://www.well.com/user/smalin/miller.html)

[5] Gouws J. and Gouws LE. (1999) Fundamentals of Software Engineering Project Management, Published by Melikon Pty. Ltd. (http://melikon.com)

OPTIMISING MAINTENANCE PERFORMANCE: AN APPLICATION OF THE THEORY OF CONSTRAINTS

Kirkwood, CK, BE Mech (1st class hons)

Team Leader Operational Performance - Asset Capability Management P/L - www.assetcapability.com.au

Abstract: The maintenance function performs a range of activities to deliver reliable equipment performance and meet business goals. This activity is described by a series of independent, yet integrated, processes that form a maintenance work management system (MWMS).

All processing systems are in some way constrained by bottlenecks. The Theory of Constraints (TOC) is a philosophy effective in optimising the performance of processing systems by working on bottlenecks.

The concept of lead and lag measures evolved from use of the Balanced Score Card. Lead measures can be levers to achieve lag measures (targets). Lead measures can be used to define bottlenecks and manage the work flow in the WMS and thereby ensure maintenance performance is improved and eventually optimised.

Direct application of TOC to managing maintenance WMS performance, using integrated lead and lag measures, derived from data captured within a CMMS, will enable an organisation to realise performance far beyond that possible by traditional approaches.

Key Words: Maintenance work management, process, system, Theory of Constraints, improved reliability.

1 DEFINING THE CHALLENGE

Maintenance has been on a journey to understand and demonstrate its' purpose in the value chain of manufacturing businesses for some time. More recently manufacturing business has come to recognise maintenance and production as two interdependent processes linked through the functional performance of the equipment. The most significance aspect of this change is that maintenance, like production, is now expected to demonstrate how it adds value to the output of the operation not be treated just as a cost centre.

Whilst maintenance personnel have always known they contributed to the plant output, being able to demonstrate how they do has been a challenge. The maintenance function has been treated as purely a cost for so long that all its management systems are a reflection of this cost driven paradigm. The new paradigm must be able to balance the cost of maintenance against the plant output it supports. Maintenance must get a new map to navigate. It must develop and manage its systems to navigate towards an optimum balance between cost and benefit and thereby achieve 'optimum maintenance performance'.

For maintenance to find this new map it need look no further than the philosophies adopted to optimise production performance. One such philosophy is the Theory of Constraints (TOC). When applied to the maintenance WMS this philosophy not only facilitates the change in paradigm required, it delivers immediate and significant benefits. These benefits include significant increases in equipment reliability, more effective work completed, reduced cost, improved morale, a common language with production and an solid foundation for equipment strategy development. Most significantly the application of TOC to the WMS delivers a systems model which facilitates the ongoing optimisation of maintenance performance.

2 UNDERSTANDING THE THEORY OF CONSTRAINTS (TOC)

All real-world systems have at least one constraint; otherwise they would be capable of infinite throughput, which is clearly impossible. TOC purports that a real-world system with more than three constraints is extremely unlikely. A major implication of this is that managing a complex system or organisation can be made both simpler and more effective by providing managers with a few specific areas on which to focus; maximising performance in the areas of key constraints, or "elevating" the constraint (making it less constraining) [1].

2.1 Five Focusing Steps

How can any generic solution have such broad applicability? It turns out that no matter what an organisation's offering is; products and/or services, the methods for most effectively and efficiently managing processes and resources are basically the same. To use another powerful analogy: just as the strength of a chain is dictated by its weakest link, the performance of any value-chain is dictated by its constraint. Recognizing this, the resulting steps to maximise the performance of a value-chain are [2]:

1. Identify the constraint.
2. Decide how to exploit the constraint.
3. Subordinate and synchronize everything else to the above decisions.
4. Elevate the performance of the constraint.
5. If in any of the above steps the constraint has shifted, go back to Step 1.

WARNING DO NOT LET INERTIA BECOME THE SYSTEM'S CONSTRAINT

Steps 1-5 can result in significant changes to the operation of the business. If this change is not carefully managed the chaos that ensues may result in the process delivering less than the potential throughput. In effect the process of changing would then be the constraint. Slowly changing, elevating constraints / subordinating non-constraints would make the system less likely to become chaotic.

3 ESTABLISHING THE FUNDAMENTALS OF THE MAINTENANCE WMS

Maintenance is viewed as a set of interdependent processes that create value by reducing operational, safety and environmental risk introduced by equipment defects. Risk is reduced by either minimising the likelihood of occurrence or by reducing the subsequent effects. Ideally the risk is eliminated through the removal of the defect all together.

Before TOC can be applied across the maintenance WMS several fundamentals need to be clarified.

1. The first is to draw a distinction between the two management systems within maintenance that form the maintenance value chain.
    a. One system deals with the application of existing equipment strategy (existing knowledge). Its processes include identification, approval, prioritisation, planning, scheduling and execution of maintenance work. (Work Orders)
    b. The other system deals with the development of new equipment strategy (new knowledge); the continuous improvement of equipment strategy, delivering new knowledge to reduce equipment defect risk. Its processes include identification, approval, prioritisation, planning, scheduling and execution improvement projects.

For the purpose of this paper only the former system is considered, the more traditional system for managing day to day maintenance work. However TOC can be applied just as effectively to the latter system but that is a story for another day.

To draw a comparison to a manufacturing process; focus on the products you need to make today not the products you are developing.

2. The second is to draw a distinction between processes and system. The maintenance processes describe how one work order would be progressed from identification to completion. The maintenance system describes how all work is managed within the constraints (Cost, Resources, Skill etc) of the system. TOC is applied at the system level.

To draw a comparison to a manufacturing process; focus on how you can achieve the desired throughput of all products instead of how you make one product.

4 BUILDING AN OPERATIONAL MODEL

To apply the TOC philosophy to the Maintenance WMS we must first define the model of the production process. The process of manufacturing completed maintenance work.

In this production process the product is maintenance jobs, referred to as "Work". There are several points of entry for "Work" into the production process and generally the later the "Work" enters the process, the more likely it is to be completed.

For this reason it is more effective to build the model by starting at the end of the production line and working backwards. The completed model is shown in figure 1.

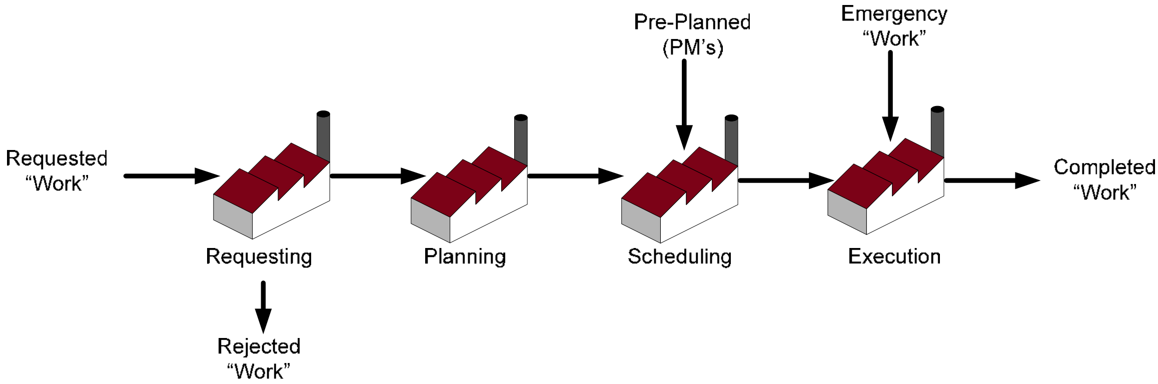


Figure 1 - Maintenance WMS Model

Remembering the product being produced is "Work", then stepping backwards from completed "Work";

- The first processing area encountered is "Execution". In this processing area "work" is done that was delivered to it from either the scheduling process or directly as unscheduled activity (breakdowns). Execution also includes the recording that "Work" was done and any identification of follow-up "Work", i.e. *History Collection*.
    - Unscheduled work should only consist of work that is so urgent that it must by-pass the planning and scheduling processes. For example if the "Work" was to be planned and scheduled then by the time it was completed the consequences it was intended to prevent would have already effected business performance.
- The second processing area is "Scheduling". In this processing area "Work" is confirmed as ready for "Execution" and then assigned within various constraints. These constraints include availability of plant, internal / external resources and budget. In some cases the "Work" is not able to be processed as parts or specific equipment / tools that are assumed to be available are not. The "Work" typically enters the "Scheduling" processing area from either the "Planning" process or as *pre-planned* work (typically PM's).
    - Pre-planned work typically comes from equipment strategies where the "Work" has been predicted and is automatically generated on a calendar basis or potentially by some event or usage based monitoring. The key difference of pre-planned from planned "work" is the assumption the specifications of the "Work" are the same each time it is processed.
- The third processing area is "Planning". In this processing area "Work" is given detailed specifications so that it can be Scheduled and Executed efficiently and effectively according to relevant company policies e.g. Safety Policy. In addition any orders for resources not available within the business are placed, such as parts or contract labour. The "Work" enters the "Planning" processing area from the "Requesting" processing area.
- The fourth and last processing area is "Requesting". In this processing area "Work" is reviewed to ensure that it is real, of a type that can be manufactured using the maintenance WMS and that an order for the same "Work" is not already in process. In addition the "Work" is tagged with a priority that effects how it is handled at the other processing areas. When this priority is determined as *emergency* the "Work" is immediately directed to "Execution" processing area. All other work enters through the Requesting" processing area. The "Work" enters the "Requesting" processing area from several sources. These sources are:
    - The need for work is recognised and a request is raised.
    - The nature of executing emergency "work" sometimes results in the quality being compromised and a need for further "Work" is requested to ensure the equipment is returned to original specifications.
    - Whilst executing Preventive Maintenance type "Work" a need for further follow-up "Work" is identified and a request is raised.

Now the WMS model is defined we can apply TOC to increase the throughput of the system and reduce the inventory or work in progress. Considering the maintenance WMS model the throughput is the amount of completed "Work" that comes out of the "Execution" processing area. The inventory is the "Work" still to be completed, stored at various processing areas.

5 APPLYING TOC TO THE MAINTENANCE WMS

Table 1 illustrates how the TOC can be applied to the Maintenance WMS by drawing on two specific examples of potential constraints (bottlenecks).

TOC step	Example 1	Example 2
Identify the constraint (bottleneck)	Planning capacity. The rate of planning "Work" cannot meet the rate of Requesting "Work". The rate of planning also cannot supply sufficient planned work to scheduling or execution.	Internal Electrical Resources Availability. In this case the "Work" requires specialised in-house skills. Currently there are not enough resources available to do all of the "Work" of this type.
Exploit the Constraint (bottleneck)	Ensure that the planner is planning as efficiently as possible. Consider reasons why planning cannot have greater throughput of planned work. Ensure that all of the information systems that support planning are effective, eg Parts Listings.	Ensure that the personnel with the specific electrical skills required are primarily focused on this "Work". Consider actions such as remove all other work from the specific electricians or allow overtime for this "Work".
Subordinate all other processes	Do not schedule or execute more "Work" than the planning rate (allowing for emergency "Work" throughput). Unfortunately it is too risky to not review incoming "Work" so the requesting process cannot be subordinated and work will build up at planning.	Subordinate the Planning and Scheduling of this type of "Work" to meet the capacity for Execution. Ensure that the preventive "Work" can still be managed as a priority over all other types.
Elevate Constraint (bottleneck)	Increase the planning capacity, either by allowing the planner more time to plan, or by providing assistance to the planners.	Increase the Execution capacity around the specific electrical skills required. There may be several ways to do this such as work overtime, remove all other work from the specific electricians or potentially recruit / contract more of this skill. Note there may be a time lag for new resources to become effective.
Check the constraint (bottleneck) has not moved	Potentially the rate of planning has now increased such that more "Work" can be planned than can be executed. Hence the bottleneck has moved to execution.	As the "Work" completed by the electricians starts to increase ensure that Planning and Scheduling can match the rate of Execution, especially in the case where the electricians' non critical skill "Work" is reallocated to others.
Manage Inertia	Slowly increase planning rate. If the amount of work planned immediately doubles then other parts of the system may be shocked, such as scheduling or supply of parts.	Steadily increase the flow of "Work" to Execution by increasing Planning and Scheduling". Continue to watch the levels of "Work" of this type as it may only be necessary to elevate this constraint for a short period.

Table 1 – Application of TOC to maintenance WMS

5.1 Introducing a Prioritisation Policy

In order to manage the flow of "Work" through the processing areas a prioritisation policy is introduced to ensure the highest risk (operational / safety / environment) "Work" is completed first and the lowest risk "Work", completed last. The prioritisation policy must also allocate priority to the types of maintenance work; Reactive, Preventive, Corrective. e.g assign emergency Reactive work as highest priority, followed by Preventive then Corrective.

If the total Work throughput goes up but low risk work is placed ahead of high or medium risk work, the impact on operational performance will not be maximised. Adherence to this prioritisation process must be achieved to maximise the effectiveness of reducing the risk due to equipment defects. The prioritisation policy will have the effect of:

1. Ensuring only emergency "Work" by-passes the Planning & Scheduling process
2. Ensuring that "Work" is processed in order of priority at each processing area.

6 MANAGING FROM LEAD MEASURES TO LAG MEASURES

For the managers of the maintenance WMS it is the lag measures, or system outputs, that are important for assessing operational performance and managing the manufacturing value chain. Lag measures for the maintenance WMS model must be defined and clearly linked to operational performance. These measures would typically be:

- Equipment Availability and Reliability to measure manufacturing plant performance.
- Cost / Unit to measure compliance to budgeted R&M cost.

These measures are known as "Lag" due to the difficult in directly influencing them. For example you cannot simply change your reliability by being more reliable. You must change the activities that you do or don't do that result in your reliability. The measures of this type of activity are the lead measures. For the maintenance WMS these lead measures are:

- Work throughput. For the whole system this is measured as completed "Work" volume. For each processing area it is the volume of work passed to the next processing area.
- Work Inventory. For the whole system this is measured as the total number of open "Work" orders. For each processing area this is the volume of "Work" waiting to be processed.
- Time to process work. For the whole system this is measured as "Cycle Time" defined as the time for work to progress through the system. For each process area this the time from receiving the work, processing it and then passing it on to the next process area.

From the two examples shown in Table 1 it can be seen that from just a single pass using the TOC process all processing areas within the maintenance WMS and consequently all of the lead measures will be affected. As the changes in lead measures occur a subsequent effect on lag measures will be observed, usually with some time delay (lag). In the examples an increase in the total throughput of completed "Work" was achieved and within a month or two the reliability increased.

Taking a systems approach of adjusting the maintenance processes, observing the system and process lead measures and the consequential effect of lagging business performance is an enormously powerful learning experience. Using this systems approach managers are able to quickly and sustainable increase the value generated by the maintenance function.

7 REAPING THE REWARDS

Significant benefits are achieved very quickly through the application of the constraining management process as well as the adherence to the risk based prioritisation of work. "Work" throughput steadily increases, reliability improves and in addition a shared understanding of the maintenance system emerges. The focus shifts from being on individual jobs to the whole system.

However to maintain the benefits the improvement must be sustained. This requires taking a holistic or systems approach across all aspects of the maintenance function e.g. Process, System, Management, and Communication. Influencing the whole function is key to the sustainability of all benefits. Like a bucket of water with a small hole, a partial dysfunction could result in all of the benefits being drained away over time.

The following provides greater detail of how these benefits are manifested.

7.1 Process Benefits

For personnel with responsibility within individual maintenance processes the benefits come in the form of more work completed through better organisation, alignment of individual activity to the system and clear goals for each process. These benefits are demonstrated by the following examples taken from the application of TOC to the maintenance WMS of several steel manufacturing plants;

a) The prioritisation of work ensures the minimum emergency "Work" occurs, hence allowing those responsible for completing emergency "Work" to be more effective. In one example a lot of breakdown work was done on standing work orders with no strategy and no planning control. As this work was reviewed for strategic purpose it allowed more work to be systematically fed to the Execution process. Typically the work on Standing orders is not well defined and allows for a wide variety of activity making it a source of inefficiency and making it difficult to assign priority.

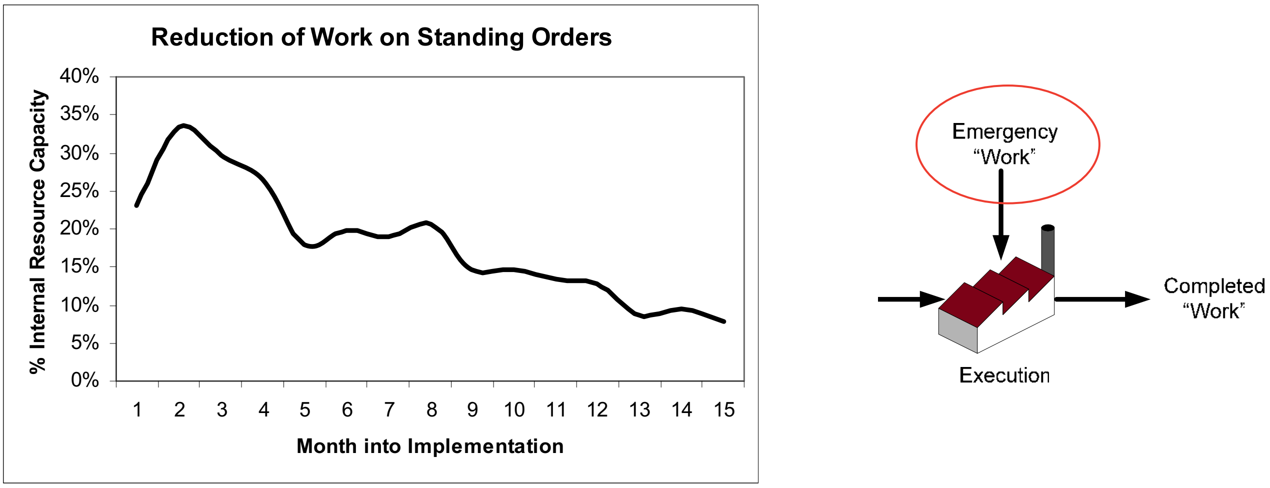


Figure 2 – Reduction of Work on Standing Orders

When the "Work" inventory at the Scheduling process is examined it typically contains a lot of duplication and redundant "Work". In two cases scheduling backlog was halved virtually over night as it contained many instances of the same preventive maintenance (PM) work order that had not been completed.

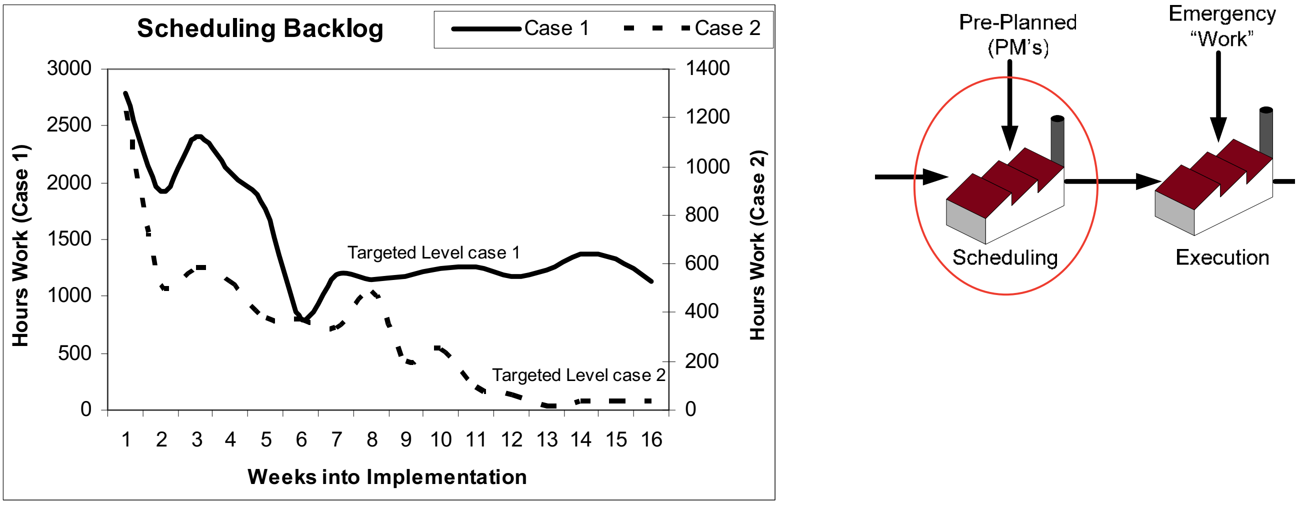
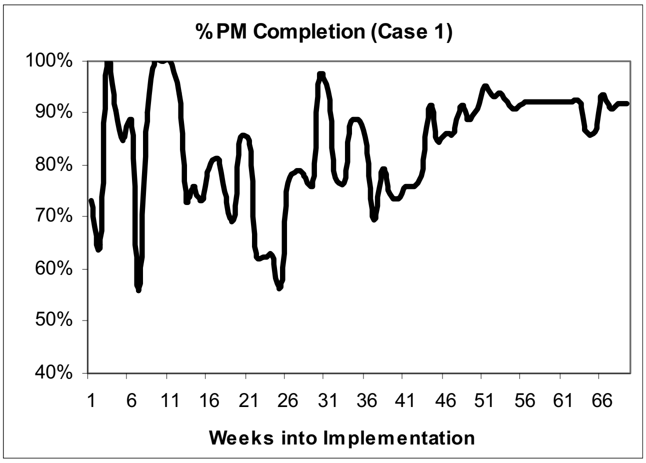


Figure 3 – Reduction of Scheduling backlog

As the scheduling backlog is work that could be done immediately this means the parts required for the "Work" are in stock. The cash tied up in the parts is only of benefit to the business when the "Work" is done. Reducing this scheduling backlog means reducing the cash tied up not working for the business. In Figure 3 the original value of the scheduling backlog was in the order of $300K.

b) Preventive Maintenance (PM) work orders are generated from an equipment strategy that is designed to minimise the risk to the manufacturing operation. Success of the equipment strategy depends not just on the completion of the PM work order but completion on time. PM work orders should be prioritised as high risk "Work" ensuring they are processed first. This will lead to a significant increase in compliance to PM due date. In the following two cases where TOC was applied, PM compliance went from an average 70% to almost 100% and the consistency (reduced variability) of performance also increased.



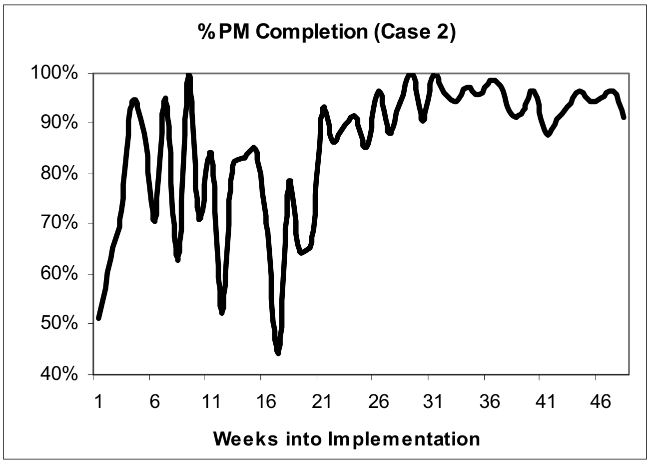


Figure 4 – Increase of PM Completion on time

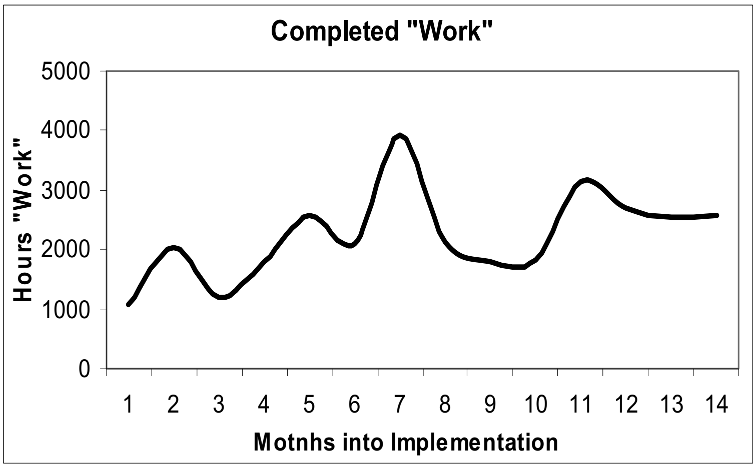
c) Using TOC the effort to manage the WMS can be more effectively distributed as subordinated processes require less effort. This effort can then be better redirected to the constrained processes or to managing the system.

- For example in one manufacturing plant the mechanical planner had no planning backlog yet the electrical planner had a significant planning backlog and was the constraint for electrical "Work". The mechanical planner started working 1 day / week on helping the electrical planner with data entry, hence allowing the electrical planner more time to plan. This effectively elevated the constraint.
- In another example the planner and supervisor were both planning and distributing all ready jobs to the tradesman. Consequently too much work was being handed out for execution each week. Matching the planning rate to the scheduling and execution rates meant only one person needed to plan. The resulting benefits were that the supervisor was able to spend more time on managing the scheduling process and the tradesmen were less frustrated as they were able to complete every job given to them.

7.2 System Benefits

As mentioned previously, for the managers of the maintenance WMS it is the lag measures, or system outputs, that are important for demonstrating their effect on operational performance. The impact of TOC on the system outputs is substantial. These benefits are demonstrated in the following examples taken from the application of TOC to maintenance WMS of several steel manufacturing plants;

d) As the constraint is exploited and the system elevated, throughput of completed "Work" increases. In one example throughput increased by approximately 60% with the same resources. Maintenance personnel also received encouraging feedback from operations who noticed a substantial increase in jobs getting done.



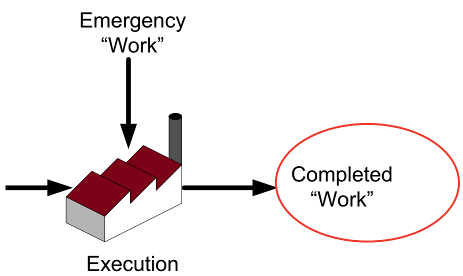
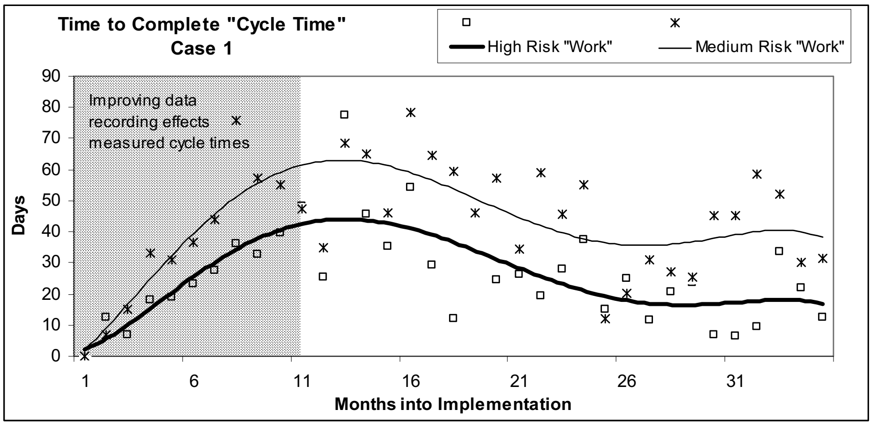
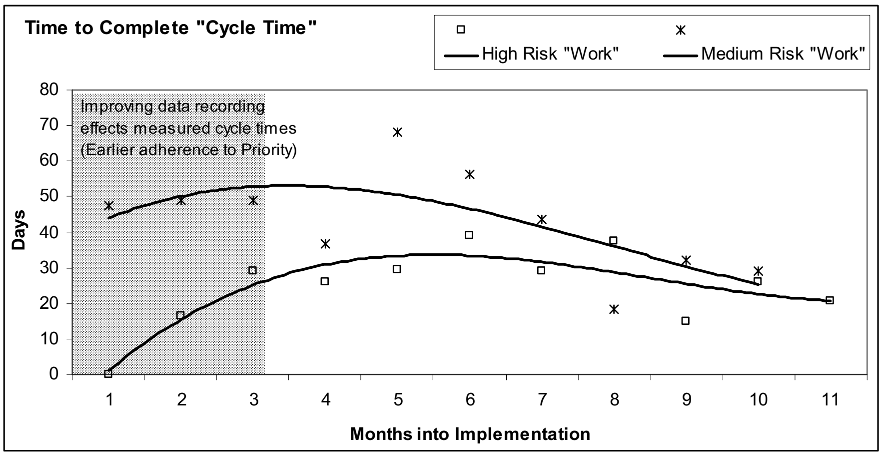


Figure 5 – Increase in completed "Work"

e) As throughput increases and the prioritisation policy is adhered to the time taken for planned "Work" to go through the maintenance WMS, known as Cycle Time, decreases. This is most significant for high and medium risk "Work". In two separate cases the cycle time appears to increase as the prioritisation policy is adhered to. Once all "Work" is completed with a recorded priority the trends show a steadily declining time to complete "Work". This effectively means the risk posed by equipment defects is removed from the system more quickly.





Figures 6a & 6b – Reduction of "Work" cycle times

f) There is an increase in the Reliability of the equipment. In two separate cases the reliability increased more than 4% over a 12 month period and then was sustained at this new level of performance.

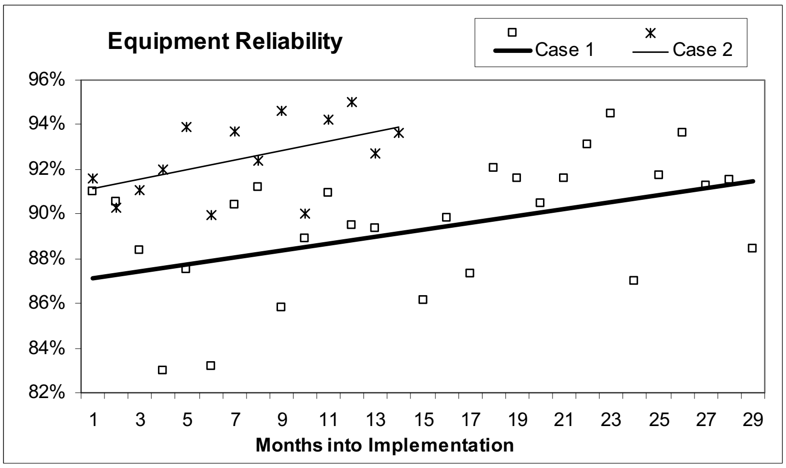


Figure 7 – Increase in equipment Reliability

7.3 Management Benefits

The old adage, "If you can't measure it you can not manage it", is just as true for maintenance. Applying TOC across the WMS and monitoring lead and lag measures is a great tool for managers to systematically link process to systems to business performance.

True commitment to measurement means that you are prepared to question every activity for its validity in relation to business performance. A key benefit of the application of TOC to the maintenance WMS is that it aligns all activity, through performance measurement, to the throughput of completed "Work". Put simply the measurement system answers the question "What is the point to doing this?" This line of thinking sparks a chain of events that ultimately reshape the culture;

1) What is our key output? What must we influence?
2) Now allow every single activity to be driven by this measure, this is performance management.
3) Now what gets done is done to serve the key output.
4) As culture is "what gets done" now the culture is aligning to the key output

Additional benefits to management are;

g) Management can determine budgets based on level of risk to the operation held in the maintenance WMS. Effectively the business is able to determine a desired cost / unit for production and convert this into an equipment reliability / availability and R&M budget. This is further explained in section 4 - ACHIEVING OPTIMUM PERFORMANCE.

h) Management benefits from the formation of teams and the alignment of performance around the whole maintenance WMS. This allows departmental division created through technical barriers such as Mechanical and Electrical knowledge to be broken down.

7.4 Communication

i) Individuals performance is aligned to a shared understanding of maintenance. For example one maintenance manager said " *I can have very quick and direct conversations with my team as to where we should focus our efforts to improve. We can stop arguing about what each measure means and start looking at the whole picture*".

j) Maintenance can explain their processes to production personnel using a picture that production are used to thinking about. Parallels between the actual plant process and the maintenance model can easily be drawn.

8 CHALLENGES TO ACHIEVING OPTIMUM PERFORMANCE

The production process, the maintenance WMS, is now much more effective at turning out complete "Work". As the constraints (Bottlenecks) are elevated the throughput of "Work" continues to increase. High risk work is completed in shorter time periods and the equipment reliability improves. This improving equipment performance feeds back to the maintenance WMS as less emergency work, allowing greater volumes of planned and scheduled work to be processed.

Effectively a self perpetuating system; improve the way you do work means less work. However after a time the increase in "Work" throughput comes at a higher incremental cost. The improvement costs more and more each time.

This concept is illustrated in Figure 8, taken from [5], where a relationship between cost of lost output and the direct maintenance cost to sustain the output is shown.

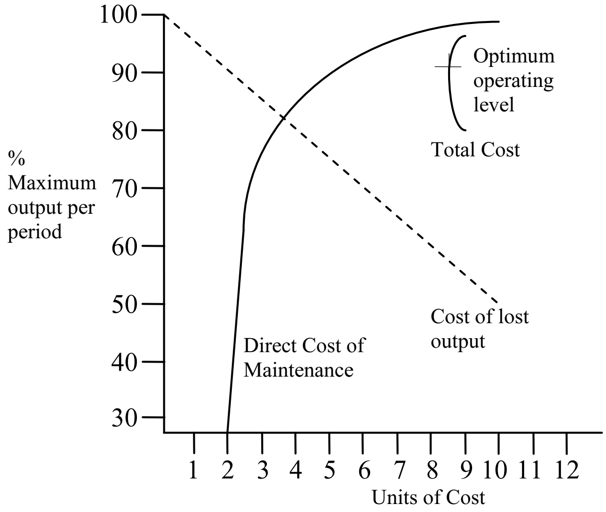


Figure 8 – Optimum Operating Performance

As the output drops from 100% the cost of lost output linearly increases as well. Similarly as output increases the direct maintenance cost to support the level of performance at first changes very little and then as 100% output is approached the direct maintenance cost rapidly increases, potentially asymptote to 100%. Hence 100% output may not be financially feasible. The resulting combination of these two effects suggests that for each manufacturing business an optimum level would exist somewhere less than 100% output.

Clearly the challenge for the maintenance manager is to achieve this optimum maintenance performance balance by monitoring plant output (Cost / Unit) together with "Work" throughput .

The maintenance WMS model is used to achieve this goal as it creates linkage between plant output (Cost / Unit) and "Work" throughput. The systems map shown in Figure 9 demonstrates how this occurs.

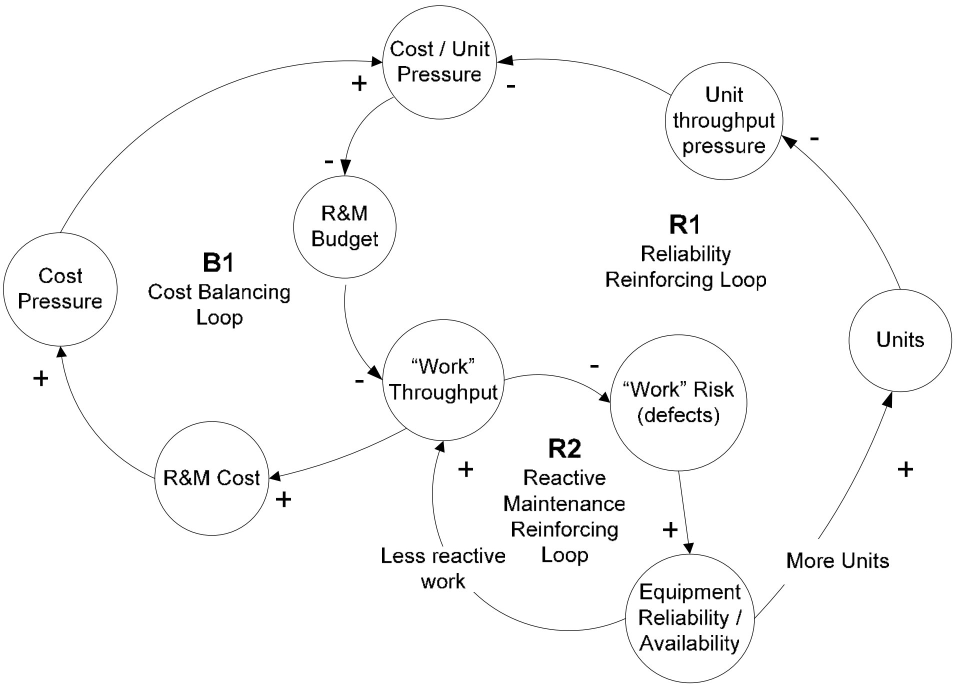


Figure 9 – Systems Map of relationship between Plant Output and Work Throughput

Starting with reliability loop R1 at "Work" throughput; as more "Work" is complete, less Risk exists which turns into improved reliability producing more units and eventually leading to less pressure on plant throughput. As the plant throughput pressure is lowered so is the cost / unit pressure lowered. This eventually leads to less pressure on budget allowing even greater "Work" throughput. This relationship tends to drive improving plant performance.

Coming to the balancing loop B1; as more "Work" is done the cost of R&M goes up and consequently puts more pressure on operating cost. This increase pressure on operating cost increases the pressure on cost / unit, which in time can lead to decreased maintenance budgets. Declining maintenance budgets means less "Work" throughput. This relationship tends to push the "Work" throughput back towards previous levels.

Finally R2, shows that a reinforcing loop feeds back from Equipment Reliability into "Work" throughput due to the level of emergency, or reactive, work. As reliability increases less breakdowns occur this allows more preventive "Work" to be completed. This relationship does not necessarily affect the amount of work done but does effect the type and the associated risk to the operation. As the type of work shifts from reactive to preventive the equipment reliability increases without associated increase in the costs of performing the work. This relationship can be used to very quickly drive increases in plant performance without substantial increases in operating cost. This is commonly referred to as "Breaking the Reactive Cycle".

In addition, in meeting the challenge of finding the optimum performance balance, it should be remembered that the WMS was defined as the application of existing equipment strategy (Section 3). This balance will be a function of the knowledge contained in the existing equipment strategies. Generating more effective strategies will elevate optimum performance.

The importance of the maintenance WMS in elevating the point of optimum maintenance performance is two fold;

a) When questioning the effectiveness of existing equipment strategy it is essential the strategy has been adhered to. If this cannot be guaranteed the root cause of the strategies ineffectiveness may be that the PM was not done on time. The maintenance WMS removes this root cause by ensuring PM compliance and reducing the time to complete high risk follow-up work.

b) When implementing new strategy the benefits are only realised when the resulting work is complete. The maintenance WMS guarantees immediate adherence to the new strategy.

9 CHANGING PARADIGMS

"A paradigm is a way of thinking; a paradigm is like a mental model. A good way of thinking about a paradigm is like it is a map. If you had an inaccurate map of the city you are in and you are trying to find a certain location what success would you have?" [4].

The plight of Australian manufacturing competing in a global market place is all too obvious. Daily media proclaims the need for numerous panaceas; innovation, improvement in managing skills, government support. [5]

For Australian manufacturing to survive it requires a change in paradigms. Maintenance can contribute to this change but it must get a new map. It must develop and manage its systems to navigate towards an optimum balance between cost and benefit and achieve 'optimum maintenance performance' and thereby show true purpose in the value chain of manufacturing businesses.

At the heart of it this is the application of TOC to the maintenance WMS. This is the map that maintenance needs to navigate its way to optimum maintenance performance where the ideal balance between maintenance cost and plant output are achieved.

The beauty of this journey is that significant benefits start to flow immediately. The immediate benefits come from the identification and exploitation of the constraints within the WMS. These include reduced risk as high priority work takes precedence, removal of obsolete and improvement work from the system, reallocation of the management effort to the constraining process and more work complete for the same cost and effort. These early wins for the maintenance personnel are crucial to sustain the motivation to change paradigms.

Within months the improvements at the work processing level start to show benefits in plant performance. This early impact on plant performance is crucial for sustaining the commitment from the whole business to change paradigms. With the commitment from the business, maintenance can then move to elevating the constraints within the WMS and start its journey towards optimal maintenance performance.

There are also many less tangible benefits for maintenance personnel. Confidence to modify failing equipment strategies knowing the root cause is not "Preventive Maintenance not done on time". The ability to explain maintenance to production using a production model removes many of the communication barriers. Setting maintenance budgets can become a question of the business managing levels of risk instead of trying to cut costs. Within maintenance barriers created by technical knowledge are broken down as everybody's focus shifts to increasing the total work throughput.

As the new paradigm solidifies it results in a changed maintenance culture. This in turn brings benefits as people are less frustrated and enjoy work more. They feel they are successfully contributing to improved business performance.

However nothing worth having comes easily. The changes that occur to the maintenance processes are essentially not that large but the changes in how the whole system is managed are. It is a big challenge to reshape the management systems and practices of the old cost focused paradigm.

10 REFERENCES

[1] http://en.wikipedia.org/wiki/Theory_of_constraints#Basic_concepts

[2] The Theory of Constraints and its Thinking Processes - A Brief Introduction to TOC, http://www.goldratt.com/toctpwhitepaper.pdf

[3] Kelly, Dr Anthony, (1998). Maintenance organisation and systems – Business Centred Maintenance. Figure reproduced from “Fig 10.2, P167”

[4] Covey SR (1991) Principle centred leadership. Free Press, Sydney

[5] Hall J Australian Financial Review, 12/1/06 “Kraft Job Cuts. Roberts P Australian Financial Review.15/1/06 “Sea of neglect..”

OPTIMISING THE MAINTENANCE FUNCTION - IT'S JUST AS MUCH ABOUT THE PEOPLE AS THE TECHNICAL SOLUTION

Dr. Brendan J Smith

Team Leader Organisational Performance Asset Capability Management P/L www.assetcapability.com.au

Abstract: It is relatively easy for engineers to think about improvement in terms of a technical solution. Typically improvements will involve the replacement of equipment or the application of a specific technique. However often the improvement, particularly large scale improvement, will not meet expectations and you will hear excuses like "…we forgot the culture" or "… we just couldn't get the people to adopt the change". Typically these words describe what most people realise; improvement or change requires consideration of both the technical and the social aspects of the system. Realising the technical side of systems change and improvement is where engineers and maintenance managers traditionally are most comfortable. This paper will deal more with the social aspects. Offering Maintenance leaders a framework and system to design and implement sustained improvement. Considering "where we are now" in order to be able to effectively plan and navigate the path to "where we need to get to".

Key Words: Leadership, Change, Systems, Assets, Maintenance, Culture

1 CHALLENGES FACING MAINTENANCE AND ASSET MANAGERS

In most developed countries, gross national product continues to increase markedly faster than population growth; this is largely the result of replacing manual operations by technological innovation [1]. Modern industrial societies are thus more and more dependent on technology to produce goods and services. The increasing complexity of technology, our increasing reliance on it, and the growing potential for businesses to incur heavy losses when their equipment is not fully operational, signals challenges ahead for manufacturing and asset managers[2].

The industrial mode that describes a stable business environment where firms compete in national markets, have long product life cycles and competitive advantages that rely on low production costs and economies-of-scale, is changing. Now markets are global and varied, competitors can access diverse labour, capital, and supply conditions and product life cycles are relatively short as new products are introduced with increasing speed [4].

The businesses that form and implement the best responses to these challenges will be those that survive to tell the story and the way that return on equipment assets is optimised, will be key to competing in the fiercely competitive global economy.

Within this market context, given that maintenance plays a major role in business success, there are a series of questions asset and maintenance managers need to be asking: (i) How will the business respond to these competitive and technological challenges? (ii) How can we manage our equipment assets in line with the business strategy and add more value? (iii) What knowledge, skills and attitudes do we need to promote and engender within our people to help us face these challenges head on?

Maintenance and asset managers have never had a better opportunity to lead, to work with the rest of the business to collectively formulate how the future should be shaped and then to marshal the resources and support to make a difference. Leadership however is a choice that individuals make to take a step forward, to show the way and to positively influence their organisation. Whilst leadership can occur in any position and at any level within an organisation, what do Asset and Maintenance managers need to consider to be better leaders and to make a difference?

2 A LEADERSHIP FRAMEWORK

The increasing need for organisations to be responsive and competitive requires maintenance leaders to harness systems and practices at the functional level that are designed to add maximum value at the business level. Dealing with both systems and practices requires a new approach to leadership, one that considers both the hard (plant) and soft (people) assets.

Crucial to this new approach is developing a shared vision or understanding of where the organisation or department is headed and what it needs to achieve. This vision is then translated through strategy, structure and culture to create behaviours and results in individuals that lead to successful functional and company performance.

Leaders can define what needs to be achieved and can set in place strategies and a structure to achieve the required outcomes but unless they motivate employees to want to achieve, then performance will always be less than what it could be. Employees are most motivated in a work environment that provides increasing challenges and support; that encourages recognition and participation and develops all employees' sense of worth and belief that they can positively impact on the performance of the company.

This work environment or culture is achieved by fostering learning, change, open communication, continual improvement and problem solving. A culture in which leaders demonstrate a commitment towards improvement initiatives and ensure that stakeholders are given the awareness, feedback and support that they need to continually learn and improve.

The framework that enables maintenance leaders to make a difference must provide a clear understanding of strategic direction, establish appropriate structure and provide for a culture that relies on employee involvement for performance success.

2.1 Think Strategically

The interaction between any organisation and its environment can be viewed as an open system where inputs, resources and assets (both tangible and intangible) are converted into outputs (finished products and services, dividends, salaries). The capabilities of the organisation describe what it can consistently do to create value for stakeholders [5] and stakeholders will include customers, employees, shareholders and regulatory bodies.

Organisations compete for customers either by offering and delivering: (1) comparable value (to competitors) at lower prices, or (2) by providing more value (than competitors) at the same or a higher price (differentiation). Competitive advantage is the ability to outperform competitors and it is usually measured by above normal profits or return on investment.

The point of strategy is to identify and exploit sources of competitive advantage. Strategy seeks to answer the question "how do we capitalise on existing resources and capabilities and build new resources and capabilities to be more competitive in the future?" Competitive advantages arise when a strategy is implemented that results in enhanced value for customers. The more difficult it is for competitors to copy and sustain this offering and the more scarce the substitutes then the more sustainable this competitive advantage will be [5,6,7].

The competitive environment requires that senior business managers think about the organisations competitive strategy. They should be asking strategic questions like: "What are our core capabilities?", "How can we use these and build new capabilities to ensure long term success?", "What are our rare and costly to imitate resources and capabilities and how can we exploit these throughout the organisation to create sustainable value for customers in our targeted market segments?" And more specifically for Maintenance and Asset managers: "how can our maintenance systems and practices be configured to enable the business to create more value for our stakeholders?"

The four generic building blocks of competitive advantage are described briefly below in Table 1 together with examples of how maintenance can be integrated into the broader business strategy and contribute to the development of competitive advantage. This contextualised or situational approach to maintenance optimisation demonstrates how Asset and Maintenance managers can provide leadership and add value in the broader business context. In short, "Plant-specific and environment specific issues should be considered when developing or improving the maintenance system." [8, pp 139].

In support of this, research has found that companies that have implemented strong quality programs (Total Quality Management or TQM); also have strong autonomous and planned maintenance systems (Total Productive Maintenance or TPM). This may be evidence of more integrated strategies being implemented as companies that continue to improve their quality (through TQM), may also need to improve their maintenance delivery system and the overall equipment performance (through TPM) [8].

This integrative approach also recognizes that the maintenance subsystem operates alongside and interacts with a myriad of other business systems. The impact of a change within one part of a maintenance system (e.g. a planned maintenance initiative) will have far reaching effects throughout the business that are not always easy to estimate (e.g. rising inventory costs, decreased staff satisfaction or improved safety performance). It is thus important for maintenance leaders to invite other stakeholders within the business to understand and participate in initiatives that may previously have been considered "maintenance only".

A prime opportunity thus exits for maintenance leaders to work with other stakeholders to explore how the maintenance function can work collaboratively to deliver better value to the business. But having convinced the business that maintenance deserves a place at the strategy table, where to from here?

2.2 Structure and Culture: Key Elements of Maintenance Design

Strategy answers the questions "what do we need to do to compete? and, what are our goals?", the organisation or department then needs to implement this strategy by designing and putting in place the appropriate structures and culture to influence the activities of people in line with the achievement of these goals.

Structure describes the formal system of task and authority relationships that influence how people coordinate their actions and use resources to achieve company goals [6]. Structure typically follows strategy meaning that the role that the maintenance department plays in the overall business strategy should influence how it is structured. Following from Table 1, if the strategy is to be quality focused then the maintenance organisation may be structured so that specific equipment strategies can be meticulously carried out by specialised trades-people, in this case the structure would include detailed formalised procedures for execution and recording of maintenance activities. If the business strategy however is to be first to market in response to customer needs, then developing rigorous equipment strategies or formalised procedures may not be warranted. The ideal structure in this case may be decentralized, lack specific planned maintenance policies, procedures and information systems and be focused on ensuring that all trades-people are highly adept at fixing any type of problem, the so called "jack of all trades". Organisational structure can be configured or designed to suit the business strategy. However structure on its' own means little, unless the culture supports it.

Table 1. Application of the generic building blocks of competitive advantage to maintenance

Building Blocks	**Features**	**Maintenance role in developing competitive advantage**
Efficiency	Efficiency is a function of inputs/outputs. Using the same level of inputs or resources to produce more output; or alternatively, producing a given level of output with less inputs is a factor that can lead to an increased ability to compete.	More efficient maintenance leads to lower costs and improved availability Planned Maintenance initiatives give management the tools to better control the maintenance spend versus equipment reliability relationship.
Quality	A product or service contains a set of attributes such as form, features, reliability, functionality, support and performance. A product is said to have superior quality when it has attributes that customers perceive to be superior to the competition	Better designed and commissioned equipment leads to increased ability to deliver quality products Better equipment maintenance systems lead to increased ability to reliably deliver superior quality products Equipment maintenance strategies developed to sustain functional performance where the defined function includes product quality as well as plant operation.
Innovation	Innovation includes advances in the types, product, production processes, management systems, organisational structure and strategy developed by a company. The ability to innovate is crucial as it is a source of uniqueness, the ability to offer something different to competitors	Maintenance people have equipment and control system expertise and intimate knowledge of the equipment; they also talk to equipment suppliers and could participate in determining what technical innovations could add the most value. Maintenance need to be structured to align equipment strategies and life cycles with the innovation strategy.
Responsiveness to Customers	Responsiveness is the ability of an organisation to identify and respond to customers needs. This leads to customers placing more value on its' products than the competition. Customer Response Time is an important aspect of responsiveness as is the ability to customise an offering to respond to the particular needs of customers or customer groups.	Maintenance need to be structured to support the responsiveness strategy e.g. if being the quickest to respond to a request for a new feature is crucial then the maintenance system must be able to respond to this need for flexibility If the key value driver is being flexible in terms of meeting a customers' "just-in-time" requirements then the maintenance system must support this.

* See [5]

The second aspect of organisational design that needs to be considered to ensure alignment with the business strategy is the culture. Culture is a more abstract but an equally important component of design; it describes the shared values and norms that influence members' behaviours and interactions with each other. The strategy is meant to clearly outline "what we need to do to compete", the structure to formalise "how we should operate to carry out the strategy to achieve our goals". The final piece of the puzzle is the culture which is "what we have learnt to do and what actually gets done". Culture is a key influence on behavior. Behavior is what gets done and people need to do the right things, otherwise the structure and strategy won't work and the goals won't be achieved. Unless the culture promotes those behaviors necessary for achieving success in the organisations' competitive environment (i.e. the strategy); the culture will become a liability for the organisation [10].

So in systems that rely on the coordination and performance of people, such as the maintenance function, improvement implies change. Change implies shared learning and sustained shared learning means a change in the culture. This sounds pretty simple, surely we just have to work out the strategy and structure and communicate this with some training and the culture will sort itself out. Well unfortunately it is not that simple and it is the lack of understanding of the complexities of culture and how to change it that often brings improvement efforts undone. Culture is actually quite complex, it operates across many levels and it is largely intangible. Figure 1 illustrates the often used onion model of organisational culture.

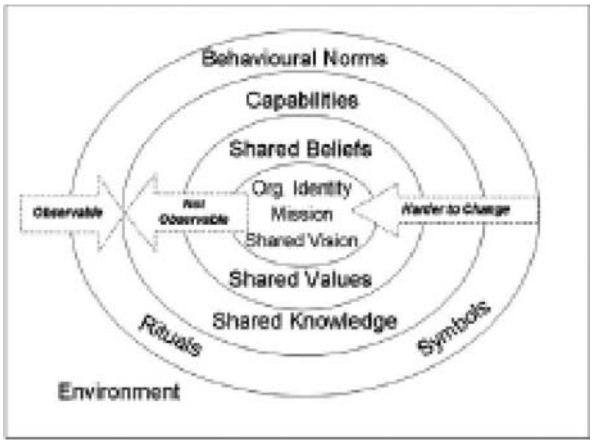


Figure 1. The Onion Model of Organisational Culture

Understanding the complexities of culture requires an appreciation that the systems that people work in are not solely technical systems. The mechanical systems that engineers design and use to manage performance are purposive; their purpose is to perform according to their design requirements. A maintenance function however contains people as well as equipment. Within this technical and social system people generate their own purposes and these do not necessarily correspond to those of the managers so these systems can have many purposes or they are purposeful. What determines an individuals' purpose in any context is their mental models or unique mixtures of beliefs and values [9]. Behaviour, which is what gets done, is the result of the purposes or intentions of employees played out in a specific environmental context.

If the culture is not considered as a part of design and all of the levels in Figure 1 are not aligned across a function or business then the individuals within the system will largely operate according to their own unique beliefs, values and motives. Typically in a maintenance function the culture and structure could reinforce a reactive work environment and the function will add little more value than just fixing breakdowns. Table 2 outlines a typical reactive maintenance culture across the levels of culture and contrasts this with a preventative maintenance culture. It can be seen that there is a significant shift required to move from reactive to preventative. It also becomes obvious why an improvement strategy like "let's just send all the planners to some training and then they can put in place the preventative maintenance system" is almost certain to fail.

Successful cultural change involves aligning all the stakeholders impacted by the system; their identity (what is my role), values (what is important), beliefs (what is the best way to maintain our equipment), knowledge and skills (know-what and know-how) and ultimately their behaviours to be consistent with the formal structure of the organisation. This type of change does not happen through individuals (even managers) working in isolation to design the ideal solution and then impose the change on others. Every key stakeholder has to want to change for the change to work, and every stakeholder has their own purposes, perspectives, and levels of understanding, that they bring to the table. Change occurs in an environment where individuals can share their concerns as stakeholders in the system and then mutually agree on courses of action to obtain the levels of performance that are required. The challenge for leaders is to be able to create the environment that encourages shared understanding and an exploration of the real issues that need to be resolved for a system to work effectively.

This implies an expanded skill set for the maintenance leader; knowledge of maintenance structures and systems needs to be matched with the skills to develop effective relationships between stakeholders and to create meaning and negotiate win-win solutions out of confusing and complex scenarios.

Many prescriptions exist to align the workforce culture with the strategy and the structure. An important ingredient in these prescriptions is 'employee involvement'.

Table 2. Clash of Cultures between Reactive and Preventative Maintenance Departments

Levels	Reactive Cultural Features	Preventative Cultural Features
Environment	Fluctuation between crisis and calmness	In control
Rituals, Symbols	The panicked reaction of the breakdown The status afforded to the most technically astute breakdown expert	Graphs on the wall The morning meeting between supervisor and tradespeople
Behavioural Norms	We play cards until we get the call Everyone drops everything and we together in a crisis Work overtime to get breakdowns fixed work	PM's are done on time and given priority Clear lines of escalation are followed for a breakdown Everyone works to Work Orders
Shared Knowledge	Everyone needs to be able to fix any problem and know all equipment systems / failure modes Short cuts to get the job done	Different types of maintenance activities and performance criteria Understand the concept of planned maintenance Understand different roles we all carry out to make the system work
Capabilities	Crisis management Able to make at least a temporary fix to all but the most serious breakdowns	Detailed and accurate planning, scheduling and execution of planned maintenance activities Able to control maintenance budget in line with optimal equipment availability and reliability
Shared Values	What is important is teamwork in a crisis Quick decision making is important in a crisis We value the time between breakdowns	Data and Measures are important Repair equipment before it breaks Process control is important Everyone understanding the preventative maintenance system is important
Shared Beliefs	Shift-men add the most value and are promoted into higher management The best way to maintain is to wait for the equipment to break, If it isn't broken, don't fix it We don't need systems and measures	Planners and schedulers add significant value and are promoted into higher management The best way to maintain is to work out and then systematically apply individual equipment strategies. We manage by systems and measures
Maintainer Identity Mission and Vision	We fix breakdowns our, mission is to do this as quickly as possible	We work together to systematically eliminate defects from an equipment system

2.3 Employee Involvement

Kearney states that employee involvement (EI) is one of the dominant elements of world-class manufacturing that can make both small and large companies competitive in the global market [11]. EI also "pervades all the other World Class Manufacturing components; TPM (Total Productive Maintenance), JIT (Just in Time), and TQM (Total Quality Management)" [12, pg 42 (parentheses added)]. Recent research also suggests a positive correlation between high levels of employee involvement and both company productivity and long-term financial performance. Manufacturers have also stated that the benefits of EI practices are employee satisfaction, quality improvement and productivity enhancement [11].

EI practices involve structuring the work environment to value employees' participation in processes such as quality improvement groups, problem solving groups, roundtable discussions, or suggestion systems. Employees are given independence in making day to day decisions and may also participate in the design of their own job, planning and scheduling of their own work and in business planning. EI relies on the manager(s) maintaining open communications with employees.

It is important to realise however that EI has its own risks and leadership challenges. A sure and well trodden path to failure comes from granting power to individuals without replacing the command and control (or reactive) structure and culture of the past with a design that provides individuals with the support to develop the knowledge, skills and confidence that they require to exercise the self judgment, discipline and leadership to succeed in their roles [13].

Despite the illusion that managers control employees and "make" them perform; employees make choices as to how diligent they are in working towards achieving their goals and the goals of the organisation. Empowered employees, the ones who make you think "if all my employees were like this, then my job would be easy", choose to do their best everyday and they frequently go above and beyond expectations. These employees are nurtured from a long term and committed effort from senior and local managers to involve employees, to provide them with the appropriate training, feedback and rewards; and by management doing what they say they will. This leadership style promotes trust and commitment amongst employees and results in psychologically empowering (or motivating) thoughts. Four primary thoughts or cognitions have been identified: meaning, competence, self-determination, and impact. In effect, to be their best, employees must believe that what they are doing is worthwhile (meaning). They must believe that they are competent to do what is expected from them (competence). They must perceive that they have choices (self-determination) and they must believe that their behavior will have some influence on what happens in their work environment (impact) [14].

Navigating the path towards an involved and empowered workforce requires the ability to lead individuals, departments and organisations through change; it could in fact be argued that this ability to foster change is the essence of leadership [15].

Both employees and managers need support to become empowered leaders in their workplace so that they in turn can assist others to change in line with the needs of the business. To achieve this effective people management practices are needed to effectively align strategy, structure and culture [16].

3 A SYSTEM TO MANAGE PEOPLE PERFORMANCE

Similar to the journey of Asset Management, People Performance Management (PPM) practices have also undergone a search for strategic relevance over the last 30 years. Whilst maintenance and engineering functions attempt to manage equipment across its' lifecycle in the best strategic interests of the business, PPM practices have focused on implementing and supporting business processes that do exactly the same thing for the people, i.e. recruitment and selection, training and development, workforce planning, performance management, career and succession planning and remuneration and reward. [16]. The analogy continues when one considers that an equipment breakdown is equivalent to a people management breakdown (a workforce strike) and preventative or proactive maintenance is akin to high levels of employee involvement and empowerment [17].

PPM practices are described by an integrated set of processes (system) that define, support, assess and reinforce employees' behaviors and results that are required for business success. The system enables maintenance leaders to implement and manage the framework described in the previous section.

PPM provides the means to set clear expectations of; the outcomes needed to meet stakeholder expectations, the activities (roles) required to make the structure work and the behaviours and values consistent with the culture and the optimal performance of roles. Individuals are appraised and given feedback, training and support so that all activity is directed towards business goal achievement. People are motivated by receiving positive feedback, training and by having personal interests, career plans and rewards aligned with business needs.

The result is a motivated, focused, competent and committed workforce, that can use work processes more effectively and efficiently to deliver better value to business stakeholders.

4 A HOLISTIC APPROACH: WHERE IS THE EVIDENCE?

It is relatively simple to measure the performance of tangible assets like equipment, using such measures as Overall Equipment Effectiveness (OEE). Measuring the performance of intangibles like people performance is far more difficult.

On a business scale there is a significant amount of research purporting the benefits from utilising the leadership framework and PPM. One research study collected data across nearly 1000 organisations on the link between their High Performance Work Practices (HPWPs) and hard measures of company performance such as employee turnover, sales revenue and profit. Companies were thus assessed on their adoption of HPWPs, which were divided into two major factors (i) employee skills and organizational structures and (ii) employee motivation. Employee skills and organizational structures describe mechanisms where employees could enhance and use their knowledge, skills, and abilities to perform their roles. More specifically job design programs, improved selection processes, formal training, quality of work life programs, quality circles, and labour-management teams were all assessed. Information-sharing programs, formal grievance procedures and profit- and gain-sharing plans help to increase the probability that employee participation were also a part of this factor. The employee motivation factor was composed of a more narrowly focused set of High Performance Work Practices designed to recognize and reinforce desired employee behaviours. These practices include using formal performance appraisals, linking those appraisals tightly with employee compensation, and focusing on employee merit in promotion decisions. It was found that the magnitude of the returns for investments in HPWPs is substantial. A one-standard-deviation increase in HPWP's yielded a relative 7.05 percent decrease in employee turnover, $27,044 increase in sales, $3,814 increase in profits and $18,641 increase in (company) market value annually on a per employee basis [19].

Further evidence comes from a case study in an Australian steel manufacturer (Australian Maintenance and Engineering Excellence Award - 2005) where the mechanical maintenance department was functioning in a well controlled preventative manner. Yet the electrical maintenance section was still deeply entrenched in a reactive culture and structure that reinforced breakdown management as the predominant mode of operation and did not believe that preventative maintenance principles could apply to electrical equipment. Electrical equipment performance was limiting the overall equipment performance of the maintenance department.

To promote a change within the electrical section sponsorship was provided from operations, production and maintenance management and other stakeholders included the electrical tradesman, a union representative and the production shift managers were engaged. Together an electrical design was formulated that outlined the ideal culture and structure required for the electrical section to function effectively alongside the mechanical section and meet the goals of all stakeholders. This shared vision that was developed through involvement of key stakeholders was then successfully implemented through consideration of both the soft (e.g. open and regular communication to teams and individuals, wide involvement, employee satisfaction) and hard (e.g. CMMS and process training, documented roles and procedures, targets were agreed around the break up of maintenance activities, openly displayed hard performance measures) aspects of change.

The result was that electrical maintenance equipment reliability improved to the point where it was no longer the bottleneck on the overall maintenance performance. In addition the new understanding and level of trust that emerged between electrical and maintenance personnel lead to the cooperative resolution of many issues that had previously been explained as the electrical sections issues. Perhaps the best evidence of cultural change comes from the electrical shift workers, who traditionally are dedicated to reactive maintenance; in the electrical section they are now spending over 30% of their time on planned and scheduled preventative maintenance activities and are key supporters of the new system

The evidence is thus starting to mount that this holistic approach works, people who recognize and respond to this will be the leaders of the future.

5 "WHERE ARE WE NOW - WHERE DO WE NEED TO GET TO"

The rapid development of technology, the ever increasing focus on customers and the ever expanding globalisation of markets are leading to the increasing need for organisations to be more responsive and competitive. Only those organisations able to respond to these challenges will succeed. The challenge for maintenance and asset managers is to collaborate with other business leaders to define, implement and nurture the structures and cultures that will enable the business to succeed. Doing this will require a new set of leadership skills; they will behave in a way that demonstrates that the people who operate and maintain the equipment assets are a key component of their success. Typically this will involve designs that encourage and support a flexible and empowering work environment.

Not to forget the importance of the technical side of improvement. Along with the need for involvement and empowerment and effective people management practices, there is also a need for a well understood system and for processes (e.g. equipment strategy) and tools (e.g. a CMMS) that operate and interact with other business systems in alignment with the business strategy. So it is a holistic approach; leadership plus systems, tools and technologies (to automate systems and manage data) plus competent, empowered, satisfied, performance focused people that will deliver plant performance to meet business goals.

It is difficult to fully understand, to bring everything together and consider all the relevant aspects of a holistic approach, particularly if you extend this to thoughts like; "where are we now" and "where do we need to get to". Organisational audit or assessment methodologies that consider such a holistic approach to improvement do exist. Models are available from The European Foundation for Quality Management, Maintenance bodies - MESA (Aust), SMRP (USA), IFRIM (EU) and The Australian Business Excellence Framework to name a few.

Maintenance and asset managers are thus encouraged to engage the support they require either from within their business or from the wider business community, and to go on a journey to become leaders and to further foster leadership throughout their organisation.

6 REFERENCES

[1] Hill T. (2005) Operations Management: Second Edition. New York: Palgrave Macmillan.

[2] Murthy DNP, Atrens A & Eccleston JA. (2002) Strategic Maintenance Management. Journal of Quality in Maintenance Engineering, Vol. 8 No. 4, 2002, pp. 287-305.

[3] Shi Y & Gregory M. (1998). International manufacturing networks—to develop global competitive capabilities Journal of Operations Management, Volume 16, Issues 2-3, 195-214

[4] Nahm AY, Vonderembse MA, Koufteros XA. (2003) The impact of organizational structure on time-based manufacturing and plant performance. Journal of Operations Management 21 281–306.

[5] Hill CW, Jones GR, Galvin P. (2005) Strategic Management: An integrated approach. Milton, Queensland. John Wiley and Sons, Australia, Ltd.

[6] Jones GR. (2004) Organizational theory, design, and change: Fourth Edition. Australia, Pearson Education

[7] Barney JB. (1995) Looking inside for competitive advantage. Academy of Management Executive, 9 (4) 49-61.

[8] McKone KE, Schroeder RG, & Cua KO. (1999) Total productive maintenance: A contextual view. Journal of Operations Management 17, 123-144.

[9] Jackson MC. (2003). Systems Thinking: Creating holism for managers. John Wiley and Sons, Australia, Ltd.

[10] Cabrera EF & Bonache J. (1999) An expert HR system for aligning organizational culture and strategy. Human Resource Planning, 22 (i1) 1-10.

[11] Kearney W. (1997) A proven receipt for success: the seven elements of world-class manufacturing, National Productivity Review, 16, pp. 67-76.

[12] McKone KE, Schroeder RG & Cua KO. (2001) The impact of total productive maintenance practices on manufacturing performance. Journal of Operations Management, 19, 39–58.

[13] Senge P, Kleiner K, Roberts C, Ross RB & Smith BJ. (1994). The fifth discipline fieldbook: Strategies and tools for building a learning organization. Nicholas Brealey Publishing: London.

[14] Robbins TL, Crino MD & Fredendall LD. (2002) An integrative model of the empowerment process. Human Resource Management Review 12, 419–443.

[15] Covey SR. (2004) The 8th Habit: From effectiveness to greatness. Simon & Schuster. Sydney.

[16] Ulrich D & Brockbank W. (2005) The Hr Value Proposition. Harvard Business School Press: Boston USA.

[17] Kirkwood CK. (2002) Personal Communication.

[18] Kirkwood CK. (2006) Optimising maintenance performance: An application of the theory of constraints. (It is anticipated that this article will appear in the ICOMS 2006 conference proceedings.)

[19] Huselid MA. (1995) The impact of human resource management practices on turnover, productivity, and corporate financial performance. Academy of Management Journal, 38(3), 635-672.

“THE END OF THE BEGINNING”

Ross Francis

Ross Francis Consulting Pty Ltd

Abstract: Plant commissioning is possibly the shortest process in the asset life cycle of major assets. It is a critical time in the life cycle and a phase that frequently gets little attention. Never before this period and never again is there such a frenzy of activity and an accumulation of excited people with differing expectations and influences.

- **What is involved in the process?**
- **When does commissioning start and when does it end?**
- **How do we manage this process it to the best advantage?**
- **Is it too late to have any significant positive impact on asset life?**
- **Can poor commissioning have an adverse impact on reliability, integrity and life?**
- **How do differing expectations, influences and constraints affect the outcomes?**

Commissioning is the time when you first see the conceptual plant in its physical form and realise what you have got, what you haven’t got and the gap between what you wanted and what you will live with. It’s a time of good news and bad news but also a time when good management can still have an influence on the ongoing asset life cycle.

Keywords: Commissioning, asset life cycle, commissioning process, reliability, asset integrity, expectations, constraints, project, construction, operations, commissioning management, operate & maintain, operability, maintainability, strategic approach, start-up curve, handover, documentation, training, schedule goals, sub-systems, commissioning phases, alignment, change management.

1 INTRODUCTION

Plant commissioning and project handover is possibly the shortest process in the asset life cycle of our assets. It is, however, a critical time and a period that can be quite confusing and difficult to manage. Never before this period and never again throughout the long life span of an asset is there such an frenzy of activity and an accumulation of excited people with differing expectations and influences.

- What is involved in the process?
- When does commissioning start and when does it end?
- How do we manage this process to the best advantage?

- Is it too late to have any significant positive impact on asset life?
- Can poor commissioning have an adverse impact on reliability, integrity and life?
- How do differing expectations, influences and constraints affect the outcomes?
- Do sound commissioning principles apply to ongoing minor projects, modifications and scheduled plant shutdowns?

Commissioning is the time when you first see the conceptual plant in its physical form and when you realise what you have got, what you haven't got and understand the gap between what you wanted and what you will live with. It is the moment of truth, a time of good news and bad news but also a time when through good management we can still have an influence on the ongoing asset life cycle.

2 COMMISSIONING & ASSET LIFE CYCLE

On a time scale Commissioning occupies a very small part of the Asset Life Cycle as shown in Figure 1 below. Depending on the industry and types of equipment, the commissioning period is often days or weeks in a total life cycle for the assets of several decades.

Generally the Asset Life Cycle is shown with the four main phases of Plan / Design, Construct / Acquire, Operate & Maintain and Dispose / Divest. Commissioning is usually regraded as part of the pre-operations phase but in fact it fills a very small fissure between the Construct / Acquire and the Operate & Maintain phases. This often means that it falls between two quite different groups (Construction & operations) even if it is managed and resourced internally in the operating organisation. Because of the short time available and its place in the big picture it is a process that is frequently under managed, under resourced, poorly planned, and given a low priority often leading to ongoing operational difficulties. Efficient and effective commissioning requires a strategic approach, clear thinking, a sound process as well as committed and dedicated management.

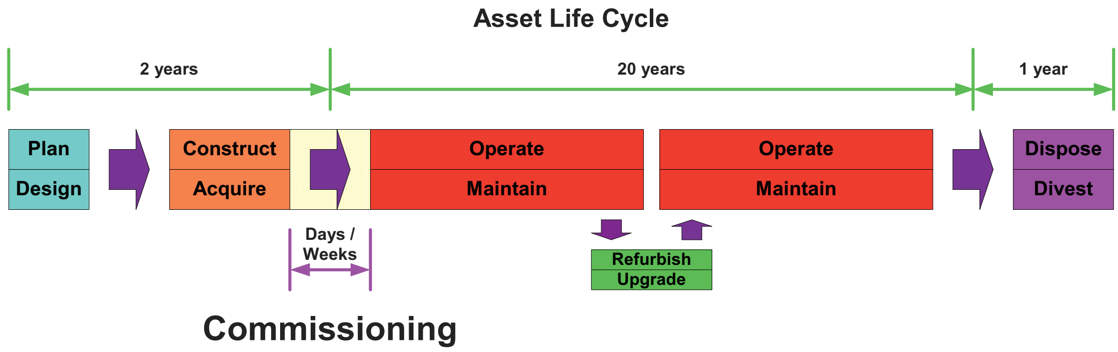


Figure 1

3 THE GOALS

- A strategic approach begins by establishing Commissioning Goals. These may include:
- Secure an efficient and quick start-up of production facility (start-up curve)
- Validate that the project construction team has installed according to the design requirements
- Ensure 100% mechanical, electrical and automation tests done before production start-up
- Clarify management processes and responsibilities before handover

- Demonstrate and document that all plant and process systems are operational
- Identify deficiencies and make corrections
- Adjust or modify equipment for best operability and maintainability
- Maintain cost, time, quality and OHS&E control to defined standards
- Phase out of construction and phase in of operations
- Complete handover documentation
- Ensure operations receive the appropriate hands on training and exposure during this time
- Progress commissioning in accordance with project requirements and long-term asset goals

4 COMMISSIONING PROCESS

Activities performed on a construction site after installation and before start-up are called ***Commissioning***. The objective of the commissioning process is to bring the plant into production with a minimum of problems. During the commissioning there is a finishing of construction, and a concentrated focus of machine vendors and a phasing in of maintenance and production activities. This complex operation in the life cycle of a facility, when it changes from the Construction / Acquire Phase to the Operate / Maintain phase, is shown in Figure 2 below.

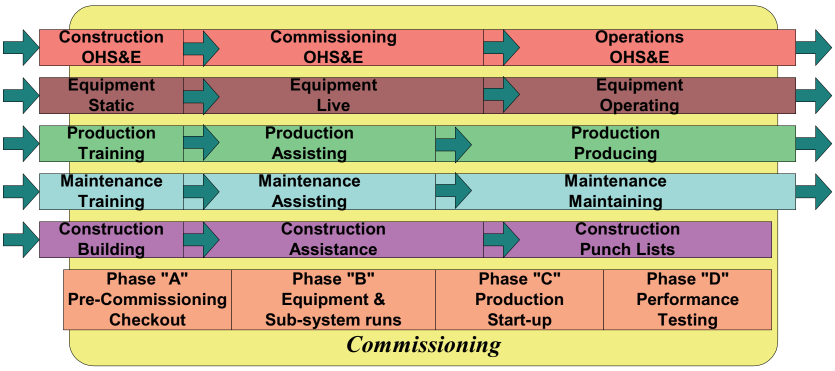


Figure 2

This picture illustrates that the roles of people involved in construction and plant operations change in nature during commissioning along with many key activities such as construction, maintenance, OHS&E. The installed equipment too changes from dormant to live, dirty to clean and untidy to orderly.

It is this phasing effect that makes it difficult to place the commissioning process in either of the main Asset Life Cycle phases on either side (Construct / Acquire and Operate / Maintain).

The commissioning process is depicted in simple terms in Figure 3 below. Although the process flow here is really concerned with the short term of the commissioning activity it can be seen from the list of inputs that the complete process of commissioning commences very early in design and is re-visited at critical stages during the Plan / Design and Construct / Acquire asset life phases.

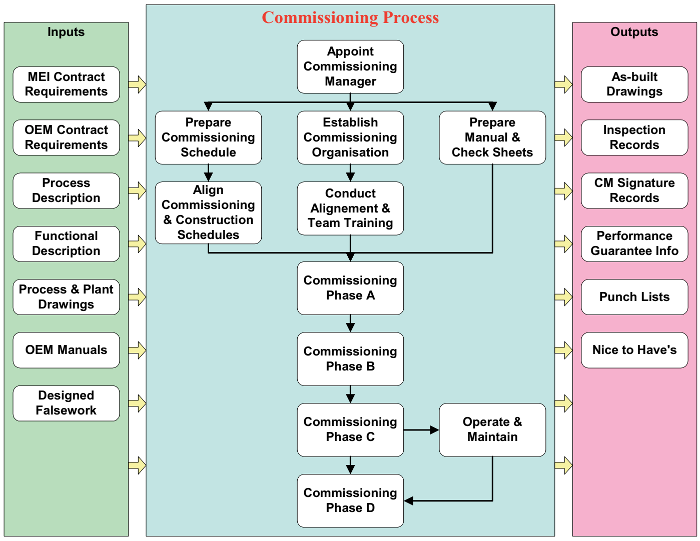


Figure 3

5 THE FOUR PHASES OF COMMISSIONING

Early planning during the construction phase and the preparation of an appropriate strategy for the plant commissioning is important to later success. This must be purpose generated to suite the circumstances and will obviously vary from project to project dependant on the plant configuration, contractual responsibilities of the parties and the requirements of the end user. A strategy for a large project for example may have, as it's key premises: -

- Divide the plant into manageable sub-systems
- Arrange the sub-systems into a logical sequence
- Assign target dates for the construction completion of each sub-systems
- Progress commissioning in pre-determined phased sequence
- Commissioned support systems (services) first followed by process systems

Having determined the overall strategy and established procedures for all parties to follow, the physical commissioning work can then proceed on a phased basis. Major phases for a process plant may include: -

- Phase A – **Installed Testing -** Normally the direct responsibility of the construction management company and / or the MEI contractors. This includes installation checks, point-to-point checks, system flushing and equipment rotation.
- Phase B **- System Checks** – The responsibility of commissioning team. Involves running equipment and conducting inspections and tests and checking operation to design. This critical phase would also be used as an opportunity for training of operators and maintenance personnel in the practical aspects of the new plant operation.
- Phase C **- Production** Dry and Wet Running. Still the responsibility of the commissioning team but a key time in the handover to operations
- Phase D **- Performance Guarantee** Testing. Carried out at appropriate times during the final stages of production runs or final operations.

A good method of illustrating and managing commissioning is to view it in four phases A, B, C & D as indicated in the Figure 4 below for a process plant (Pulp & Paper) commissioning:

5.1.1.1.1.1 5.1.1.1.1.2 PHASE	A **Completion of Construction**	A **Installed Testing**	B **Water Runs**	B **Stock Runs**	C & D **Production & Start-up**
RESPONSIBILITY	Construction Suppliers	Construction Suppliers	Commissioning Construction Suppliers Operations	Commissioning Construction Suppliers Operations	Operations Suppliers
MEDIUM	None	None	Water	Stock	Stock
MECHANICAL, PIPING & CIVIL	Complete erection Grouting Cleaning Supports Painting Punch lists	Vendor checks Safety, guards Visual Internal inspections Drive belts Alignment Lubrication Hydrostatic	System flushing Isolations on Supplier tests Service tests Functional tests	Operation Supplier tests	Acceptance Handover Optimisation Performance & Guarantee tests Full system start-up
ELECTRICAL	Installation complete Control circuits Power ready Point to point	Emergency stops MCC activated Rotational tests Dry interlocks Group starts	Interlocks Groups starts Motor loads Validate process description	Motor loads Interlocks Groups starts	Automatic ops
INSTRUMENTS & DCS	Installation complete Field checks complete DCS Installed Loop sheet verified Valves stroked Punch list	Loop tests Functional tests Dry interlocks Initial settings Site acceptance tests DCS complete	Interlocks Group starts Loop tuning Calibrations Validate process description	Consistency Calibrations Loop tuning	Consistency Calibrations Loop tuning Start up

Figure 4

6 ORGANISATION

All stakeholders must feel ownership, take responsibility and facilitate the transition of ownership. This requires a sound, fluid, composite structure and the appointment of a suitably experienced "independent" Commissioning Manager. The manager should be a person experienced in project work as well as operations and have sufficient status to be able to manage all of the parties involved. A generic organisation chart is shown in Figure 5 below.

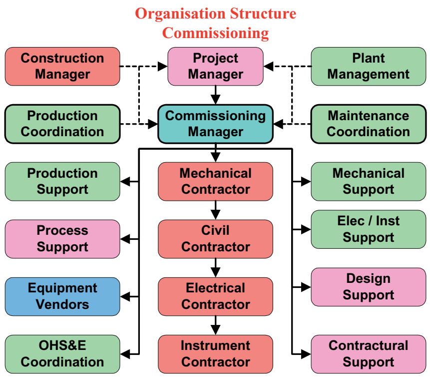


Figure 5

The commissioning manager key responsibilities include:

- Alignment of all parties involved in Commissioning of each major plant section
- Coordinate contractor commissioning assistance
- Manage changes required to process and equipment and as necessary call for revised HAZOP
- Manage the affixing of "Ready For Testing" tags and the issue of "Notice of Energisation" and all applicable OH&S requirements
- Conduct commissioning progress review forums
- Preparation of commissioning action lists
- Problem resolution directly or via appropriate referral
- Manage the handover from construction to operations
- Manage handover package (documentation) delivery
- Ensure environmental requirements are covered
- Ensure all parties work together for the same agreed goals

An early involvement is required for the commissioning manager so he can:

- Development of a comprehensive procedures manual incorporating all relevant existing operational and safety procedures
- Determine plant boundaries covered by the commissioning procedures
- Define and communicate the commissioning manager's responsibilities
- Coordinate the development of commissioning check sheets and books
- Define live equipment tagging procedures for those areas of plant not covered by the existing danger tag procedure.
- Clear determination of all relevant contracts clauses covering any aspect of commissioning
- Ensure new confined space and other procedures
- Determine required sequencing of runs, flushing, start-up etc
- Coordinate commissioning responsibilities
- Determine shift requirements
- Determine communication requirements (e.g. radios)
- Coordinate the preparation of a commissioning schedule

There are many players in the commissioning melting pot including:

- Contractors - mechanical
- Contractors - civil

- Contractors - electrical
- Contractors – instrument
- Project management & staff
- Design team representatives
- Equipment OEM & vendor representatives
- Government authorities
- Production personnel
- Maintenance personnel
- Training staff
- Auditors
- Senior management
- Public relations
- Visitors & spectators

The Commissioning Manager will coordinate and manage these individuals and groups who all come to the scene with widely differing priorities and backgrounds.

An early involvement of maintenance and production personnel, preferable those involved in design. This is important from a training perspective and also to get the early hands on experience necessary to continue into operations

6.1 COMMISSIONING ARRANGEMENTS

There is a range of normal project and operational management activities that continue through commissioning, however a few need emphasis during commissioning and a concentrated effort from the Commissioning Manager:

6.1.1 Alignment

Even if an alignment was conducted for the project at commencement it is important to have a specific alignment for the commissioning phase. This is an excellent way for the Commissioning Manager to kick off the process. It is important for all to understand how this time is seen through the eyes of others and to align the team to one set of goals and a common schedule

6.1.2 Communication

Extra effort on person-to-person communication and information circulation are essential to maintain a clear focus and to maintain efficiency across all work groups. Schedules are important as are punch list etc but time is too short to miss a daily all-in communication where issues and priorities can be raised

Daily stand up communication meetings to keep all stakeholders informed, decide on the daily priorities in accordance with the schedule and to allocate small teams or individuals to resolve issues that arise

6.1.3 OHS&E

A critical topic for the commissioning phase and as a minimum must be discussed at the daily meetings. Apart from the work complexity, housekeeping challenges this is a time of significant change from project to operational conditions.

Tagging of equipment to ensure that visually all site personnel are aware that equipment is changing from dormant to active. The concept is to have site equipment tags to inform all parties of the condition and inspection stage of every piece of equipment and to have an overall picture recorded progressively in a master folder.

6.1.4 Training

Production and maintenance training would have commenced well prior to this time. It is important however to capitalise on the unique opportunity to understand the new equipment from personal learning and technical aspects, data collection, instrument settings performance measurement and vibration signatures etc.

A wide range of personnel can get experience and learning by being involved with equipment activities such as equipment numbering and tagging, pipeline identification, checking installation and equipment to drawing and inspecting for operability and maintainability. Allocating appropriate technicians and operators to OEM's is critical.

6.1.5 Procedures Manual

Although each project is unique a procedures manual should include the following key sections: - introduction / purpose / structure, safety, isolations (testing & operations), confined space, environment, commissioning procedures, commissioning schedule, organisation, responsibilities & relationships, communication channels, handover, regulatory requirements, general procedures, vendor requirements, contacts etc.

The procedures manual can be combined with the master handover records file required to capture all inspection, test and defect information.

6.1.6 Schedule

It is important to have a schedule for all commissioning activities specifically construction, commissioning and production. The schedule will allow for inspections of the installed equipment, actioning of high priority punch list items and dovetail in with the completion of construction schedules that may continue on into the commencement of commissioning

The Commissioning Manager will manage the process to ensure that the work is carried out systematically and in the correct sequence. The sequence is important to keep personnel focussed and also to allow for system operational constraints that require some equipment and systems to be functional before others. Generally services need to be commissioned first

6.1.7 Change Management

Commissioning is the most comprehensive change management exercise that will be undertaken by a manufacturing organisation. Changes will impact on, or be influenced by, all areas of business including:

- People - recruitment, training, technical and operational safety knowledge, and communication.
- Process - design, regulatory, process descriptions, safe operating limits, and abnormal operations.
- Assets - physical changes, design standards, regulatory requirements, CMMS, accounts, procedures, records, maintenance regimes.
- OH&S - legislative (e.g. MHF) requirements, Safe working environment, dangerous goods
- Environment- approvals are gained and records modified
- Business - objectives, economic or on non-economic justification
- Procedural - codes, standards (internal and external)
- External - quality of feedstock, delivery arrangements, transfer equipment, upstream process changes, services supply, land use change, addition of plant or changes to nearby facilities.

It is important to have change management processes in place. Clearly some of the change process commences very early in the project, however it is during commissioning that it all becomes critical.

It is important to have a review and approval / rejection process for plant changes (modifications) that are raised. As a minimum they should be categorised into must do, do at some time and "nice to have's" to be listed for review and justification. Whilst not forgetting the project imperatives of cost, time, quality and scope any defects or unsatisfactory operability and maintainability aspects will most likely remain with the assets for many years.

7 WHERE DO WE GO WRONG?

7.1 People & Processes

Many of the problems in commissioning emanate from operations and project management not giving the commissioning processes and the commissioning team sufficient importance, authority and attention. This results in organisational and people aspects not being effectively controlled. Some of the key aspects are:

7.1.1 Alignment & Involvement

Not having an alignment or not have a complete (all stakeholders) effective alignment severely impacts on the success of the commissioning process.

7.1.2 Involvement

Not having the right people available or dedicated to the project at the right time. It is important to involve production and maintenance people from very early design. Unfortunately this involvement is not always seen as a priority at that time. This extends to not making a commitment in design phase or then again at commissioning to dedicate operations maintenance and production personnel to the project. Commissioning is a time where a large number of tasks must be completed. It is also the best time for training. Operations people involved at this stage get an unparalleled learning through experience opportunity that cannot be equalled late in the asset life. There is also the opportunity to influence the equipment and process that is finally handed over.

7.1.3 Communication

Commissioning is not a controlled ongoing situation that most are used to in industry and neither is it a controlled steady change as in a project, rather it is a constantly changing environment with errors, omissions, oversights, defects and physical changes coming to the forefront. There can never be enough communication, particularly face-to-face and particularly between combatants that may normally communicate only infrequently. Special effort is required to achieve an integrated effort from all of the participants and this area requires alert commissioning management.

7.1.4 Training & Understanding

The introduction of bad practices through low involvement from the early project conception and poor training prior to and during commissioning. Irreparable damage can be caused to machinery by operators and maintainers not running the plant the way it was designed to be operated. This leads to frustration, breakdown, and incorrect operations leading to very early requests for plant modifications. "Read the instructions first" still applies.

Formal training is essential there is no substitute for a wide range of plant people developing an in-depth understanding of the equipment and production process design and how it is meant to operate.

7.1.5 Process & Procedures

Operations management often assume that commissioning will occur as a natural part of the construction project, and what they miss will be picked up once it is operational. They frequently fail to realise that there are some very specific processes that must be put in place and managed. An organisation is only as effective as its processes. The process is best presented to the complete team in the form of a Procedures manual.

7.1.6 Schedule

Most participants enter the commissioning phase with their own departmental schedules that they have been working to, particularly project schedules, or the schedules that are in front of them, notably the forward production forecasts and associated budget promises.

7.1.7 Change Management

Regrettably most plants commence operations with inadequately prepared work teams, particularly from the aspect of understanding the changes and the differences to the old plant they are so familiar with. Engineering and maintenance documentations and technical data frequently is slow in being delivered and inadequate. This is compounded by the operations people not being mobilised in sufficient time before operation commences.

Inadequate data and documentation emanates from not tying down the documentation requirements at the design phase and not specifying documentation and schedule requirements at negotiation and purchase time.

7.2 Engineering & Equipment

Failure to set up the commissioning team adequately will result in short cuts being taken often resulting in damage to machinery occurring, often unnoticed at the time. Below are a few examples of minor problems that can have lasting deleterious effects on the plant, subject operations people to unnecessary personal risk and increase the ongoing cost base:

7.2.1 Design

- Valves that need to be operated frequently being place above floor access level and no fixed means of access.
- Poor lighting in dark corners
- Pumps positioned so that piping prevents crane access for maintenance
- Long equipment aligned incorrectly for 2 hoist crane lifts
- Poor access to key pieces of equipment resulting in excessive manual efforts for maintenance and operations

7.2.2 Construction

- Grouting installed to ensure water is trapped and diverted in under sole plates – unfinished work
- Hurried appearance painting in place of protective coating
- Improper line flushing and screening with foreign material remaining in the systems
- Pipe supports not in place
- Forgetting to grease, not recognising that some remaining orifices are in fact grease points or over greasing

7.2.3 Inspection & testing

- Sealing water systems forgotten and not operational
- Plastic, wood, oil dry in the wrong places so that it can enter the process system

- Leaving an oil filler cap of the hydraulic system and undoing all the pre-commissioning filtration and reducing the equipment life.
- Making assumptions on equipment condition before inspection and testing is complete
- Failure to run in equipment such as seals

Significant design problems that do not have an immediate effect on the plant operation is unlikely to be corrected at commissioning and will have a detrimental effect on life cycle operations thereafter. Other items, such as valve access, can be corrected during construction and during the frenzy of project completion at commissioning. This can only occur by having some budget provision and people with appropriate experience "on the spot" to pick up these items and initiate the corrective actions.

8 WHAT CAN WE DO

Is commissioning the end of the beginning or the beginning of the end? It could be the start of premature asset degradation if errors occur and are not noticed during commissioning, or if poor start-up practices are adopted, even if unwittingly.

Good start-up and long asset life can be assured by having experienced maintenance and production people involved in a project from a very early design stage. This needs to be seen as a priority from an asset life cycle perspective. Although the vast majority of asset life cycle costs occur during the long (20+ years) operate and maintain phase the opportunity to influence this in terms of maintainability, reliability, operability, integrity and life is at critical times during the design phase. There is a further very small but important window of opportunity to further this cause during the phase of commissioning. Regardless of how many drawings we study, even three dimensional walk throughs, there is no substitute for seeing the new plant in real life and making final corrections within the project constraints.

Sound commissioning principles certainly apply to ongoing minor projects, modifications and scheduled plant shutdowns. In many industries shutdowns are infrequent and of relatively long durations (weeks) resulting in a natural tendency to employ project management techniques and sound commissioning and handover processes. Other industries, however, have more frequent and smaller duration shuts (hours / days) and the pressure to produce often results in the sound commissioning principles being bypassed. Unfortunately this has the same effects as project commissioning in that it will affect the performance and life of the assets. The same sound principles need to be applied to these operational aspects including the appointment of a Shutdown Manager (cf. Commissioning Manager) for all shuts. Even the smallest of "projects" require the same change management and commissioning principles and practices.

The most important thing we can do is to educate all stakeholders, particularly senior management, in the importance of life cycle considerations and the importance of each phase. Project specialists are notoriously poor at developing any understanding of the life cycle. Their focus is different for obvious reasons and centres on the project time and cost budget. It is often said that quality is a given but this is only so if the correct input is applied. In all aspects it is important to take a long-term strategic asset view to achieve best benefit from the commissioning process.

VALUE DRIVEN PREVENTIVE MAINTENANCE ACTIVITIES USING QUADRANT ANALYSIS

Deryk Anderson

Oniqua Enterprise Analytics, Brisbane, Queensland, Australia

Abstract: An organization's maintenance strategy may consist of many thousands, even tens or hundreds of thousands of preventive maintenance activities. These activities, properly executed, determine the success of the preventive maintenance program. The management of large volumes of preventive maintenance activities is a challenge for maintenance practitioners. Maintenance activities do not exist as a single entity within the work management system, and it is often difficult to associate performance indicators with individual activities. Additionally the lack of a methodology for segregating and managing activities as a whole has not been well documented. Quadrant Analysis can be used to visualize and segregate maintenance activities based on simple measures of business risk, cost and performance. This paper explores the application of quadrant analysis to the management of preventive maintenance activities. Actual data is demonstrated and shows how quadrant analysis can be used to drive strategies for data collection, determine causal analysis strategies, and apply risk based maintenance analysis. Examples and case studies from Australian and overseas organisations are presented in this paper.

Key Words: Maintenance, Strategy, ERP, EAM, CMMS, PM, Quadrant, Boston, Matrix, Analysis, Activities, Risk, Cost, Analytics

1 INTRODUCTION

Asset intensive organisations in the private and public sectors rely on assets to provide products and services. The ability of these organisations to provide desired service levels, function safely (with regard to personnel, public and the environment), and economically or profitably is dependent on the effective maintenance of these assets. The assurance of asset capability [1] to meet these business objectives is the role of the maintenance function.

Preventive Maintenance designed to avoid unplanned component failure [2] is an essential part of a maintenance program. The preventive maintenance program consists of a multitude of time based activities that are directed at assets and their components. The development and continuous improvement, or management, of these activities should require a significant proportion of the maintenance management effort.

The challenge for maintenance is to effectively manage large volumes of Preventive Maintenance activities. This paper explores the application of quadrant analysis to the segregation and development of policies for the ongoing management of preventive maintenance activities.

2 PREVENTIVE MAINTENANCE

In this paper the term Preventive Maintenance [3] refers to any activity that is designed to:

- Predict the onset of component failure,
- Detect a failure before it has an impact on the asset function,
- Repair or replace asset before failure occurs.

Based on available literature, it is most likely that the development of failure prevention and management policies for most facilities will be based on a combination of:

- Operational research;
- Risk based maintenance analysis;
- Quality approach;
- Non-Risk Based Approach.
- Event and situation analysis.

These categories are described in the following sections.

Operational Research

Operational research is concerned with defining quantitative solutions to problems of Maintenance Management. The approach is to define a maintenance problem in logical terms and then express the problem as a mathematical relationship. The solution to the maintenance problem is then derived from the solution to the mathematical relationship. Operational research in maintenance is typified by work by Jardine [4], which describes models for maintenance strategy decision making based on the analysis of historical data

Risk Based Maintenance Analysis

Risk based analysis refers to the development of failure prevention and management policies based on an understanding of the consequence and probability (hence risk) and modes of failure of an item. Methodologies that describe the development of failure prevention and management policies are numerous and often amount to variations of a similar fundamental approach.

This type of approach is typified by inductive techniques such as Reliability Centred Maintenance [5]. Other examples of these techniques are Failure Modes, Effects and Criticality Analysis (FMECA) [6].

Quality Approach

The Quality Approach refers to methodologies for developing maintenance strategy influenced by Quality Management approaches, which emerged from Japan in the 1940's. These systems provided the foundation for the evolution of quality maintenance systems which were initially aimed at the application of quality tools to maintenance related problems. The quality school of maintenance includes Total Productive Maintenance (TPM) described by Nakajima [7], "Total Productive Maintenance (TPM) is productive maintenance carried out by all employees through small group activity. In TPM the machine operator is responsible for the maintenance of the machine, as well as its operation.

Non-Risk Based Maintenance Analysis

Non-Risk based maintenance analysis refers to the development of failure prevention and management policies without understanding the consequence and probability of failure of an item. These methodologies are classified as:

- Manufacturer's recommendations
- Personal Experience
- Past practice

Event and situation analysis

Failure prevention and management policies are developed to produce certain outcomes described by the objectives of maintenance. These objectives make specific assumptions about organisational and operational requirements and conditions. The development of maintenance strategy is also carried out with the use of the best available knowledge about item reliability, local work practices, and external (e.g. legal and regulatory) requirements. With time, these elements may change, and the maintenance strategy must change to remain relevant to the original maintenance objective (i.e. it must be a dynamic and evolving strategy).

A facility can be made up of thousands of individual assets, tens or hundreds of thousands of individual components [8] and tens or hundreds of thousands of individual maintenance activities. The development and management of maintenance activities for facilities of this magnitude is an enormous undertaking in terms of effort and data management. If a maintenance analyst was to do nothing else but review the performance of 10,000 maintenance activities over the course of one year – the application of simple mathematics shows that they would have just over nine minutes per activity to devote to the task. There is a need to be able to segregate maintenance activities for control purposes.

2.1 Control of Maintenance Activities

In many businesses, Preventive Maintenance activities have been established over time with little technical discipline supporting the decision process. This has resulted in Preventive Maintenance activities that:

- Are ineffective in detecting the onset of failure,
- Duplicate the effort of other preventive activities,
- Are missing for critical failures.

Control of the maintenance activities involves modifying the description or timing of the activity as changes are detected in the performance of assets or the organisational environment in which the activity is applied. This control can be required for the following reasons:

- The original assumption on which the maintenance strategy was developed is now wrong (e.g. incorrect assumption about failure rates or failure consequences)
- Changes to equipment design or parts supply

- Changes to the organisational or operational environment or requirements
- Changes to external environment or requirements.

The effective control of maintenance activities is challenged by the volume of activity data as well as a lack of classification of maintenance activities and a framework for their segregation and management.

3 THE NATURE OF PREVENTIVE MAINTENANCE ACTIVITIES

A preventive maintenance activity is a specific instruction to perform a maintenance action to address a specific failure mode or range of failure modes for a particular component. Simple examples (additional information would be required to represent a quality instruction) of preventive maintenance activities are:

- Check Gearbox Oil Level every 3 Months;
- Replace Oil Filter every 1000 Hours;
- Measure brake pad thickness every 5,000 kilometres.

Preventive Maintenance has a few apparent features or attributes: An activity to be performed, a frequency or interval (the timing) at which the activity is performed and material and labour resources required to perform the activity. The maintenance planning process generally assumes that all maintenance activities are equally important, and it is difficult to discriminate between activities just by the attributes stored in work management systems. In reality there are many other attributes of a maintenance activity. For the purpose of this paper, examples of these attributes are limited to:

- Criticality – The impact of the particular failure mode of a component that the maintenance activity is designed to prevent;
- Cost – The labour, material, services, overhead cost of the maintenance activity;
- Frequency – The frequency or interval of the maintenance activity;
- Repair Time – The average repair time of the failure mode of a component that the maintenance activity is designed to prevent;

These attributes are explored in the following sections.

3.1 Criticality

Criticality describes the impact of a failure in terms of its effect on the safety and profitability of the business. An example of a means of classifying equipment is demonstrated in Table 1.

Table 1 – Assets Criticality Rating [3]

Code	Criticality	Business Impact
S	Safety / Environmental Critical	Safety / Environmental
10	Major impact on output of all machines/lines, contributing to long factory outages. Orders lost on all products.	Total Loss of Production
9	Major impact on output of one or more machines/lines, contributing to long machine/line outages. Orders on specific products are lost.	Total Loss of Production
8	Significant impact on output of all machines/lines. Inability to meet customer order quantities by delivery dates.	Partial Loss of Production
7	Significant impact on output of one or more machines/lines. Unable to meet specific product order quantities by delivery dates.	Partial Loss of Production
6	Minor impact on output of one machine/line. Production can be rescheduled. Unable to meet some customer orders for a specific product by delivery dates.	Partial Loss of Production
5	Loss of assets with 100% capacity backup. Failure of backup assets will affect ability to meet customer order quantities by delivery dates.	Negligible Impact
4	Loss of assets with some contingency or buffer storage. Significant impact on output of single machine/line (with no affect on ability to meet orders).	Negligible Impact
3	Loss of assets with significant contingency or buffer storage. Significant impact on output of single machine/line (with no affect on ability to meet orders).	Negligible Impact
2	Minor impact on output of single machine/line (with no affect on ability to meet orders) e.g. short reduction in production capacity.	Negligible Impact
1	No affect on output of any machine or process line within repair time.	Negligible Impact

The application of criticality ratings to individual maintenance activities in a case study from [3] is shown in Figure 1. This chart shows that a high proportion of activities, by cost, were directed at the prevention of "Negligible Impact" failures or failures with low levels of "Loss of Production".

The application of the criticality attribute provides one way to discriminate between maintenance activities. There may be an opportunity to reduce maintenance effort on equipment failure modes with negligible impact.

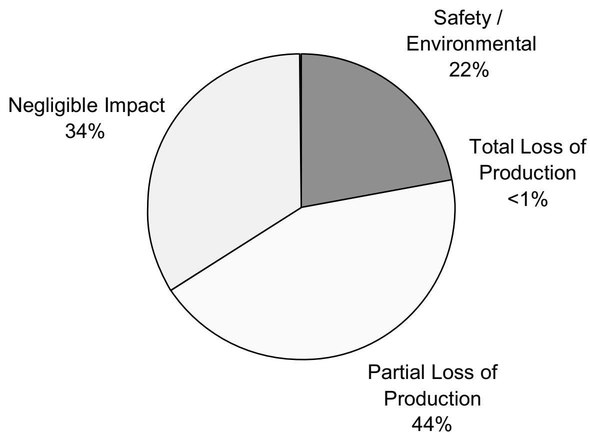


Figure 1 - Preventive Maintenance Activity against Business Impact

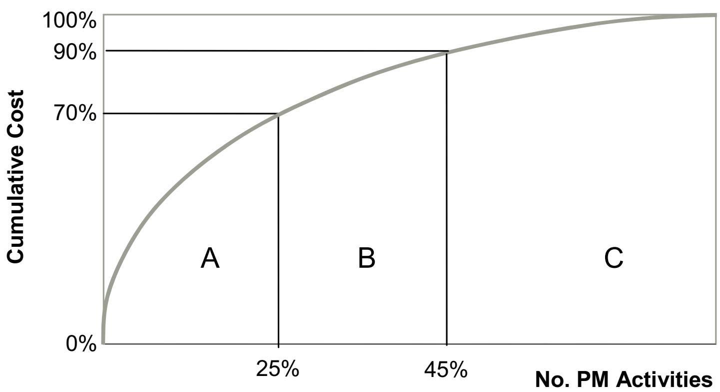


Figure 2 – ABC Analysis [3]

3.2 Cost

The total cost of performing a maintenance activity is the summation of labour, materials and services costs. One means of discriminating maintenance activities by cost is to use an ABC analysis - a common method for classifying inventories aimed at identifying critical inventory items. ABC classification is based on the Pareto Principal (Vilfredo Pareto (1848 – 1923)), that a relatively few items account for the majority of any given result. ABC classification argues that a relatively small number of inventory items account for the major part of the total inventory value. When applied to Preventive Maintenance activities a similar pattern is seen.

Figure 2 shows the results of applying ABC analysis to the Preventive Maintenance activities in the case study. The graph of Figure 2 shows that:

- 70% of Preventive Maintenance Costs were incurred by 25% of the Preventive Maintenance activities
- Just under half of the Preventive Maintenance activities accounted for 90% of Preventive Maintenance costs

ABC Analysis is another way of discriminating between maintenance activities. Combined with the previous measure of criticality further categorisations are possible – e.g. High Cost (A Category) activities on Negligible Impact failure modes etc.

3.3 Other Attributes

There are many other attributes available to categorise maintenance activities. Examples of these are frequency and repair time.

An example of the allocation of activities by cost to frequency is shown in Figure 3. The visibility of activities in this way identifies a challenge or opportunity – that of potential over maintaining. Particularly when combined with the information from the previous attributes.

The combination of information provides a point for managing maintenance activities. How do I manage my activities that are directed at negligible impact modes of failure that have a high cost and a high frequency?

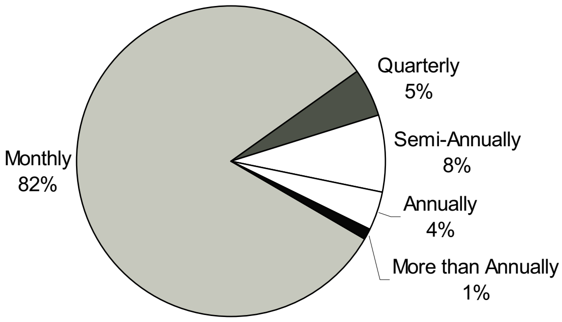


Figure 3 - Preventive Maintenance Activity Costs by Frequency [3]

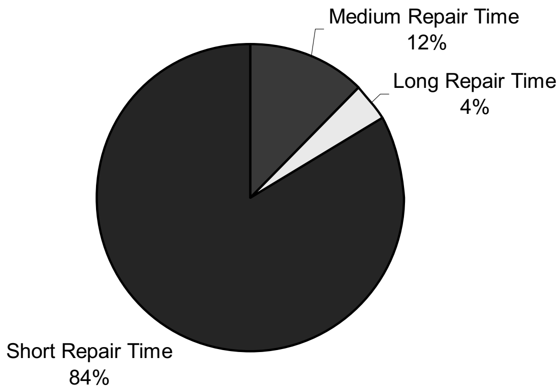


Figure 4 - Preventive Maintenance Activity Costs by Frequency (Source – Author)

A further and final example of an attribute is Repair Time for the failure mode shown in Figure 4. A repair time category can be applied to a component or failure on the basis of historical experience. For example, a Short Repair Time category may be 0 to 2 hours, while a medium repair time is 2 to 6 hours and a Long Repair Time greater than 6 hours.

A combination of attributes can create a profile for a maintenance activity that can be used to identify the activity for a specific course of management action. In this way there are no longer tens or hundreds of thousands of activities to be managed, but clusters of maintenance activities to be managed in a specific way.

There is potential for multiple attributes to become unwieldy if used to profile activities. The attributes described above give 180 different combinations or potential profiles for maintenance activities. The remainder of this paper considers the use of two attributes and four combinations to explain the features and benefits of the profiling approach. Quadrant Analysis is used for visualisation purposes.

4 QUADRANT ANALYSIS

Quadrant Analysis is a methodology of graphical representation and visualisation of any number of entities by two attributes. Quadrant analysis is viewed as a scatter plot of two variables on a set of axes to understand relationships between these variables.

An example of such an application of quadrant analysis is the Boston matrix [11] or Boston square analysis developed by the Boston Consulting Group during the 1970's [12] to plan market strategies. In the model of Figure 5, an organization's products and services were analysed in terms of Market Growth (percentage change in sales in a product or service category) and Market Share (ratio of an organization's sales in a product or service category to the total value of sales in that category).

An organization's products and services are positioned according to their perceived values and ranked within the various quadrants. These products and services are then categorised and managed accordingly. Cash Cows are used to finance other initiatives; problem children are monitored – hoping that they will eventually become stars. Dogs may be dumped.

Quadrant analysis has been extended to general problem analysis including risk assessment [9, 10]. The quadrant analysis approach is applied to the problem of maintenance activity management.

The Quadrant Analysis can be applied to the problem of maintenance activity analysis. Two possible attributes are identified in the content of Section 3, namely Criticality and Cost. A matrix with these attributes is shown in Figure 6.

4.1 Application of Quadrant Analysis to Maintenance Activities

The quadrants are described as:

Trivial – Maintenance Activities in this category have a low cost (Category B or C from Figure 2) and a low criticality (a score of 5 or lower in Table 1).

Excessive – Maintenance Activities in this category have a high cost (Category A from Figure 2) and a low criticality (a score of 5 or lower in Table 1).

Critical – Maintenance Activities in this category have a low cost (Category B or C from Figure 2) and a high criticality (a score of 6 or higher in Table 1).

Strategic - Maintenance Activities in this category have a high cost (Category A from Figure 2) and a high criticality (a score of 6 or higher in Table 1).

The naming of the quadrants is somewhat arbitrary. It is the management strategies for the Maintenance Activities in the quadrants that is the key point of the segregation. Management strategies for maintenance activities are argued on the basis of achieving the objectives of maintenance and opportunities to reduce direct maintenance costs.

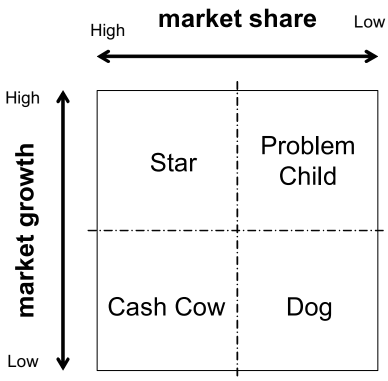


Figure 5 – Boston Matrix

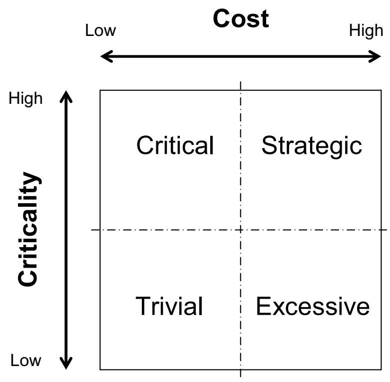


Figure 6 – Quadrant Analysis Applied to Maintenance Activities

4.2 Management Policies for Maintenance Activities

Management policies for the various segment of maintenance activities are influenced by observations on the features of these segments. These are discussed below.

Trivial maintenance activities have a low cost and a low criticality. Typically these activities take little effort to execute. The impact on the organization of not performing these activities, or working with less than optimal activities or timing is low. The effort required to make changes to these maintenance activities may be significant.

Excessive maintenance activities have a high cost and a low criticality. These activities may cost more to execute over the life cycle of a component than the opportunity cost associated with component or asset failure.

Critical maintenance activities have a low cost and a high criticality. These activities are significant because of the impact of failure. There is little opportunity to reduce maintenance costs by analysing these failures.

Strategic maintenance activities have a high cost and a high criticality. These activities are significant because of the impact of failure and the cost and effort associated with their execution.

Management policies for the various categories of maintenance activities are proposed in Table 2.

The management policies of Table 2 are considered to be a work in progress. While the content of the management policy for each segment is arguable, the diversity of the management policy is demonstrated.

5 APPLICATION OF QUADRANT ANALYSIS

To demonstrate the application of quadrant analysis to maintenance activity segregation, over 7,000 activities for a manufacturing facility were analysed. This facility was unusually appropriate for such an analysis because:

1. Recent work had been conducted to isolate individual maintenance activities from the Computerised Maintenance Management System;
2. A criticality assessment of assets and components was available to associate with individual maintenance activities.

Equipment criticality was allocated to the maintenance activities as shown in the scale of Table 1. A random factor was applied to provide a spread of points within the score band for visualisation purposes. The cost of the maintenance activity was converted to an equivalent annual cost based on the cost of the single activity annualised by its frequency. An A, B, C analysis of maintenance activity costs was conducted as indicated in Figure 2. Costs were adjusted with the A, B and C bands for visualisation purposes.

A scatter plot of the cost and criticality data was created. The number of maintenance activities in each of the four quadrants was counted. The results of this analysis are shown in Figure 7.

These results analysed by numbers show that relatively few maintenance activities were assessed as Strategic or Critical for the facility studied. Figure 7 shows that in terms of numbers (a count of maintenance activities in each segment):

Table 2 – Activity Segment Management Policies

TRIVIAL	- Perform activities as they arise (they do not take much time or effort); - Manage by exception (monitor metrics or KPIs by exception); - Delegate execution of these maintenance activities to other resources (e.g. contract or operations); - Consider activity redesign or elimination in the case of off-line maintenance activities (i.e. where downtime is required to perform the activity – or the maintenance activity is intrusive); - Pay a low level of attention to these maintenance activities (there are more important activities to be managed).
EXCESSIVE	- Consider task elimination based on economics; - Consider increasing maintenance intervals
CRITICAL	- Monitor compliance of these maintenance activities; - Monitor measures relating to cost of the activities on an exception basis; - Analyse unplanned failures related to these activities using causal analysis techniques; - Review the effectiveness of the maintenance activity on a periodic (say annual) basis; - Ensure that accurate maintenance records are kept; - Consider the use of risk based methodologies such as FMEA or RCM to improve task effectiveness, particularly on highly critical activities.
STRATEGIC	- Monitor compliance of these maintenance activities; - Actively apply risk based methodologies such as FMEA or RCM to improve task effectiveness and reduce costs; - Revisit the risk assessment associated with the maintenance activity on a periodic (say annual) basis; - Analyse unplanned failures related to these activities using causal analysis techniques; - Ensure that accurate maintenance records are kept.

- 5% of maintenance activities fall into the Strategic segment.
- 18% of maintenance activities fall into the Critical segment
- 23% of activities fall into the Excessive segment
- 54% of activities fall into the Trivial segment

The analysis of segment numbers show that relatively few maintenance activities are indicated as strategic and critical to the organisation.

If this is analysed by cost (sum the cost of the maintenance activities within each segment, the results show:

- 14% of total cost of maintenance activities falls into the Strategic segment.
- 7% of total cost of maintenance activities fall into the Critical segment
- 56% of total cost of maintenance activities fall into the Excessive segment
- 22 % of activities fall into the Trivial segment

The analysis of segment costs indicates the opportunity to challenge excessive maintenance activities.

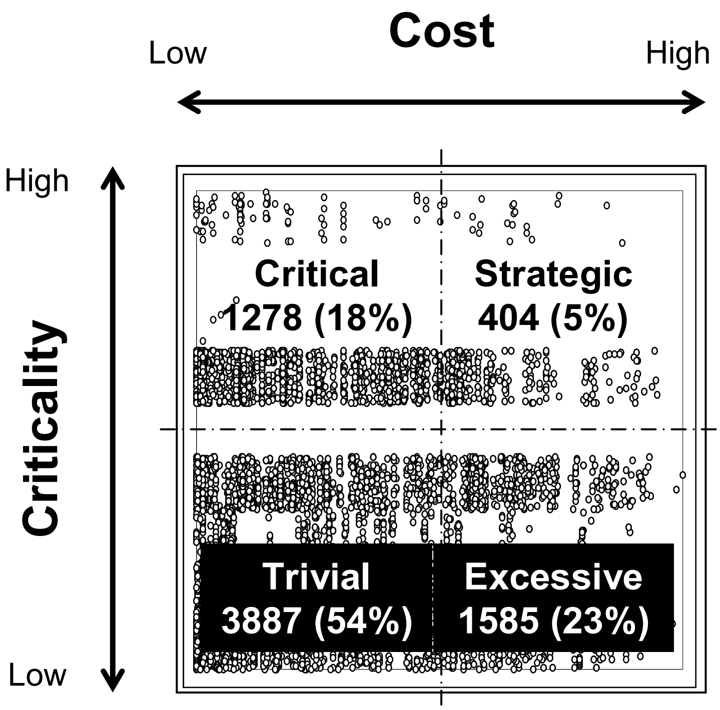


Figure 7 – Scatter Plot for Quadrant Analysis Application

The immediate benefit of maintenance activity segregation lies in the reduction of excessive activity. A study by Anderson [3] reports a reduction in the direct cost of Preventive Maintenance Activities by frequency reduction. This reduction was reported to be 40% of the total cost of Preventive Maintenance for the facility studied. Significant benefits are evident in maintenance activity segregation.

The results of the single application of quadrant analysis are not conclusive. There is insufficient analysis across sites to conclude a pattern of population of segments as a general rule. The following is indicated by these results:

1. There is variability in the criticality of maintenance activities within a facility;
2. There is variability in the cost of maintenance activities within a facility;
3. There is variability in the distribution of maintenance activities by cost and criticality;
4. Segmentation of maintenance activities indicates that different management strategies are justified for the various segments identified.

Additional work at different sites is warranted to further explore the application of quadrant analysis to the management of maintenance activities.

Technical analysis is challenged by:

- Lack of facilities with a complete set of maintenance activities identified as individual entities;
- Lack of facilities where a rigorous and dependable criticality analysis has been performed across the entire site

It s hoped that this paper will challenge organisations to consider the applicability and benefits of segmentation analysis of maintenance activities to facilitate access to additional data for analysis purposes.

6 CONCLUSION

There is a challenge for organisations to effectively manage large volumes of Preventive Maintenance activities. Quadrant analysis indicates that valid maintenance activity management strategies could be determined on the basis of the attributes of the maintenance activities.

Variability has been demonstrated in the application of quadrant analysis to a single facility. While this study is not conclusive, applicability of Quadrant Analysis to the segmentation of maintenance activities is indicated. It has been further demonstrated that management policies based on segmentation could be valid.

Additional application of quadrant analysis to the problem of maintenance activity analysis is warranted.

7 REFERENCES

[1] Murray M et. al., Maintenance Engineering Society of Australia Capability Assurance: A Generic Model of Maintenance, The Maintenance Engineering Society Of Australia, 1995, Retrieved from http://www.mesa.org.au/engineering.html

[2] Kelly A, Maintenance and Its Management, Conference Communication, 1989, Page 29

[3] Anderson D, Reducing the Cost of Preventive Maintenance, Maintenance Journal, Volume 15; No 4 – October 2002

[4] Jardine, A. K. S. (1973). Maintenance, replacement, and reliability. New York, Wiley.

[5] Moubray J, Reliability Centred Maintenance, Butterworth-Heinemann, 2nd Edition

[6] IEC Technical Committee No 56: Reliability and Maintainability, CEI ICE 812, Analysis techniques for system reliability – Procedure for Failure Modes and Effect Analysis (FMEA), Bureau Central de la Commission Electronique Internationale, First Edition 1985

[7] Nakajima, S. (1988). Introduction to TPM: total productive maintenance. Cambridge, Mass., Productivity Press.

[8] Gruver D, Using Pareto Analysis to Focus Reliability Improvements, 2004, Retrieved from http://www.mt-online.com/articles/0105pareto.cfm

[9] Lowy A, Hood P. (2004) The Power of the 2 x 2 Matrix : Using 2x2 Thinking to Solve Business Problems and Make Better Decisions. San Francisco, CA: Jossey Bass Business and Management Series.

[10] Nicholls J, The MCC decision matrix: a tool for applying strategic logic to everyday activity Management Decision Journal, Aug 1995 Volume: 33 Issue: 6 Page: 4 - 10 MCB UP Ltd

[11] Business Policy in Action, Volume 31, Issue 6 of Management Decision, by J Peters, MCB University Press

[12] http://en.wikipedia.org/wiki/Boston_matrix.

DEVELOPMENT OF A STRATEGIC ASSET MANAGEMENT FRAMEWORK

Ankur Maheshwari

Senior Consultant – Engineering and Asset Management
Norman, Disney & Young, Level 10, 200 St. Georges Terrace, Perth, Western Australia 6000, Australia

Abstract: Success of an asset intensive organisation is based on adoption of high quality asset management practices and systems. A strategic asset management framework (SAMF), ties an asset management organisation from end to end, and is targeted at improving efficiencies, effectiveness and overall performance whilst meeting the business objectives. A SAMF relies on three critical components – People, Process and Technology (Tools and Systems). Appropriate level of maturity is required in all the three aspects to ensure consistency and sustainability of the asset management outcomes. The development and maintenance of a SAMF is a continuous process and needs to be executed over the entire life cycle of an organisation. A sound approach has to be followed to ensure that an appropriate framework on asset management exists and is practiced and maintained by the organisation. Apart from a thorough understanding of the business objectives, a SAMF ensures harmonious operation of an organisation (consistency) and continual improvement (sustainability). This ultimately leads an organisation a step closer to achieving business excellence. The paper defines a strategic asset management framework and discusses steps involved in the development and implementation of such a framework. The approach presented in this paper can be applied to any asset management organisation.

Key Words: Asset, Strategy, Asset Management, Framework

1 STRATEGIC ASSET MANAGEMENT

The term asset management has differing meanings to different organisations. The definition of asset management may vary even within the same organisation. To some, asset management is just a means to an end, whilst to others it is the core mission. Asset management exists in one form or other in all aspects of life – be it managing one's real estate assets or industrial assets. The term 'strategic asset management' is proposed instead of the more loosely used term – 'asset management'.

There are various definitions to the strategic asset management. These can be found in numerous strategic asset management plans developed by building, university, utility, industrial and other infrastructure organisations. Each of these plans has defined the strategic asset management based on their organisations drivers and the needs. One of these definitions [1] is – "Strategic Asset Management can be defined as the planned alignment of physical assets with service demand. It is achieved by the systematic management of all decision-making processes taken throughout the life of the asset." Most of the other definitions closely align with this definition. These definitions assume a certain maturity level of the organisation, whilst may not ideally provide an environment of a complete assessment of the framework, process maturity and integration systems. This does not mean to say that there is any deficiency in this definition or associated management plans.

For the purpose of this paper we define strategic asset management as – "A process of developing, creating, maintaining and disposing assets through a complex series of interlinked well-defined processes that are continually improved, over the life cycle of an organisation, with an aim of achieving the objectives of the organisation." This definition takes asset management beyond mere operation and maintenance, and incorporates the creation of new assets as well. Traditionally asset management is only related to managing here and now assets – but this is not adequate in terms of ensuring long-term viability of the organisation. Strategic asset management is therefore beyond what is being done or how it is being done.

Just like parts of a body need honing and ongoing management to ensure sustainability, asset management processes also require continual development. Strategic asset management also involves this continual improvement of the processes, by monitoring and measurement. In a nutshell, strategic asset management matches the organisations assets (physical, human and resource) to its strategic directions. It develops the relationship between the strategic directions and the operational works.

2 STRATEGIC ASSET MANAGEMENT FRAMEWORK (SAMF)

One of the definitions on Strategic Asset Management Framework (SAMF) [2] "Strategic Asset Management Framework (SAMF) is an integrated policy strategy developed to improve asset management and capital investment across the State public sector. The SAMF is designed to ensure that the Government can continue to successfully manage its capital program while meeting its financial targets."

The Strategic Asset Management Framework as discussed in this paper is considered a more enhanced version of this definition. For the purpose of this paper we define "Strategic asset management framework (SAMF) as a tool that ties an asset management organisation's entire business operations from end to end. The framework forms the foundation on which the entire organisation operates." The key objectives of the framework are to ensure that –

- The objectives of the organisations stakeholders (shareholders, customers, regulators, employees, contractors and the community) are met.
- The performance of the organisation is continually improved.
- Long-term viability and sustainability of the organisation is ensured.
- The organisation is advancing towards a business excellence framework.

The discussion below presents a very fundamental view of the SAMF.

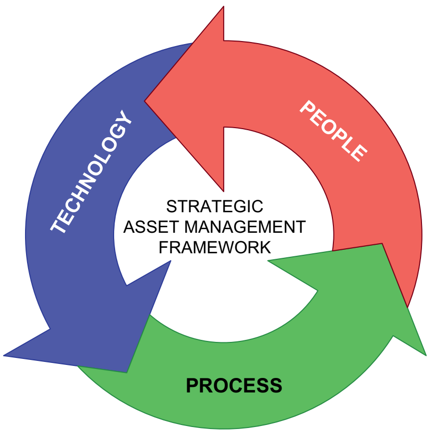


Figure 1 - SAMF Concept

The SAMF ties the three key links of the organisation namely – people, process and technology (see figure 1) and on this integration resides the delivery systems of the entire organisation.

The balance between the people, process and technology is achieved through an adequate understanding of the three links, and ensuring that an appropriate framework exists to retain the bond between these links.

People – The people aspect of the asset management framework is one of the most critical. From this comes the power of an organisation to 'think' and 'act'. It is the people of an organisation that conduct the process formulation and improvement, by applying the technology (tools and systems). Having said this, it is vital to recognise that, an organisation should render its processes and systems matured enough so that it could sustain itself even when the people turn-around is high.

Process – This is the second link of the SAMF. An organisation should endeavour to formalise all its key processes. This should be undertaken to drive value and ensure long-term sustainability. The formalisation of processes does not tend to limit the continual improvement of the organisation or enforce any form of bureaucracy. It provides a common meaning and direction to its internal and external customers, whilst presenting an orderly and matured image of the organisation. Formalisation of process also enables a framework for formal measurement of the performance of these processes and evaluation as to when the time is opportune enough to modify the processes or retain the good performing ones.

Technology – The third link of the SAMF is 'technology'. This can also be called as 'tools' and/or 'systems'. This link is made up of the tools and systems that enable the processes of the organisation. The technology arm binds the people and process links and is vital to consistency and sustainability while ensuring governance and control. The various systems deployed by an organisation need to be thoroughly integrated to ensure a smooth transition between the individual processes and deliver maximum efficiencies and effectiveness. The systems link is also the source of the critical information on which people of an organisation rely to make their decisions. Asset Management Information System (AMIS) is the key system within this link. Growing importance of the AMIS around the globe and its increasing application and maturity levels are example of the criticality of this link.

3 THE NEED AND BENEFITS

SAMF is a best practice approach to management that is utilised by high performing asset management organisations. The framework exists in one form or the other in varying degrees of maturity in each organisation, however its formal development and enhancement is what distinguishes high performing organisations from the rest. It is important to acknowledge that:

- an organisation with a SAMF does not guarantee success; and
- an organisation that is succeeding does not necessarily have a formal SAMF.

Whilst the above holds true, development and implementation of the SAMF provides a foundation for success, consistency, sustainability and credibility within an organisation. Depending on the operating environment of an organisation, the drivers for a SAMF may vary from organisation to organisation. Typically, the drivers include one or more of the following:

- consistency of application of business processes, including business sustainability;

- direction and guidance to staff (and contractors/ service providers);
- development of a culture of continual improvement;
- greater immunity from staff turn-around;
- efficient utilisation of scarce resources and appropriate allocation of resources (by prioritisation);
- minimisation of waste;
- enhanced productivity (by identification of poor performing processes);
- increased efficiency and effectiveness;
- greater returns from investment;
- critical input into strategic business planning processes;
- transparency in business operations;
- greater customer and community satisfaction;
- tool for regulatory compliance and demonstrating effectiveness of business operations.

Whilst the above are some of the many drivers for adopting a SAMF, there could be other factors that could act as catalysts to this process. These could include – scarcity of skilled resources, customer and regulator assurances and company image as a sound operating business etc.

3.1 The Benefits

As discussed above, the drivers and related benefits from a SAMF vary and it is up to an organisation's staff and management to harness the potential of this framework. Some benefits that a SAMF is known to deliver include –

- Cohesiveness of the processes (that may be operating in isolation) resulting in a transparent, integrated and commercially focused end-to-end management process.
- Consistency in application of processes and systems.
- Enhanced business image.
- Increased staff morale.
- Enhanced customer, regulatory and community confidence.
- Strengthening of inputs to the strategic planning process.
- Clear directions and objectives to management of assets.
- Sharpening the focus on organisation's risk appetite.
- Better management of organisations risk portfolio.
- More streamlined business functions and processes, leads to a more profitable commercial outcome.
- Greater efficiencies from within the business and enhanced effectiveness.
- Better positioning of the business to change and its adaptability to adjust to changes in the operating environment.
- Increased business sustainability and viability.

4 DEVELOPMENT OF A STRATEGIC ASSET MANAGEMENT FRAMEWORK (SAMF)

The complexity of the process of development of SAMF should not be underestimated. It may appear to be a simple concept, but in practice it requires careful planning, management and adequate steering to enable its success. There is no rules book to the development of the SAMF. The International Infrastructure Management Manual (IIMM) [3] is an accepted industry 'best practice' guide to asset management and has been known to be a reference document adopted by regulators and many other industry participants. This manual develops some basic concepts upon which an organisation can develop sound asset management fundamentals. Whilst these principles can be adopted in development of SAMF, it is to be acknowledged, that an organisation over the time develops its own useful and practical approach to asset management. Various factors need to be considered and analysed in this process.

This paper presents a methodology that could be applied by organisations. The most appropriate methodology has to be decided based on a critical assessment of the present state of operations. Whilst presenting the various options (where they are known to exist), it is assumed that the organisations do not have an existing formal SAMF.

An organisation could embark on the development of the SAMF, regardless of its age. However, the younger organisations (with some years of operational experience) are more ideally placed to development of a SAMF. This is assuming that the organisation has had a holistic experience of all aspects of its internal and external operating environment. As [4] puts it "In the early stages of evolution, companies that have adopted the strategic asset management business model see it first as a catalyst for organisational change and cost reduction." For other organisations, the case will more be of development with continual refinement until the most appropriate framework is developed. Nevertheless, it is important to realise that the framework is not static and needs continual review and refinement as operating environment changes and organisations performance/ performance driver changes. An organisation that has undergone a major structural review, a change of ownership, a major merger, change in regulatory environment, change in supplier/ resource environment, can also be treated as a "young" organisation and has a greater chance of seamless introduction of a SAMF. For older organisations, (with minimal recent changes), the road to introduction of the framework may be more complex and will require deployment of more complex 'change management' techniques than those that may be required for younger organisations. This is based on an assumption that older organisations have established traditional processes that will be hard to break or change. However, this may not hold true for all older organisations.

4.1 Methodology

The methodology proposed for the development of SAMF is based on the following key steps –

- Knowledge of the business.
- Clear definition of the business and associated processes.
- Development of an 'end-to-end' process model of the business.
- Definition of each process – including necessary dissection of the process into elements.
- Defining the process maturity levels and scoring system.
- Establishing the measurement framework.
- Assessment of maturity levels of existing process (and elements).
- Agreement to the targeted maturity levels for each process (and its elements).
- Development of action plans (to bridge the identified gap between targeted maturity levels and existing maturity levels) including associated expenditure budgets – for implementation.
- Prioritisation of the identified actions.
- Development of processes to ensure sustainability of performance of each process.
- Formal delivery of the entire framework.

5 STEPS IN DEVELOPMENT OF THE SAMF

The various steps for development and implementation of the framework are shown in figure 2. Each step is discussed in turn below.

5.1 Preparation

The preparatory phase in the development of the framework requires the following key steps –

- Appointment of a champion to lead the development of the framework – this should ideally be a senior management who has the overall accountability for process improvement, and a thorough understanding of the entire business process.
- Formation of a team of professionals (including process owners) who will be in charge of developing and implementing the framework. The team could be split between – the resources that participate in development (strategic and tactical management) of the framework and the resources that participate in implementation (tactical and operational management). A good overview of the 'total asset management process' is presented in [3].

- An understanding of the organisations objectives and targets (including its obligations and commitments to staff, customers, regulators and community).
- A process view of the organisation (input – output cycle).
- Organisation chart (present and foreseeable future) – this should include a critical evaluation of the enablers and blockers to the asset management framework.
- An understanding of the organisation's suppliers, service providers and contractors – this should include high-level information on the objectives of these service providers. More emphasis should be placed in understanding how the supplier's objectives relate to the organisations own objectives.

The above information needs to be collected and critically analysed. The analysis of this information will result into formation of a strategic plan for development and delivery of the SAMF. This analysis will identify any likely pit falls (blockers) to the framework and the enablers to the framework. The strategy should be to utilise the enablers while ensuring that the blockers do not have negative influence. This can be achieved by adopting appropriate change management techniques involving education, controls and (if necessary) removal of the blockers. It is important to note that the framework development and its implementation will need to be established as a 'non-negotiable' agenda within the organisation. *Observation – Whilst an organisation will have existing inherent strengths, it needs to en-cash on these strengths to improve in the areas where it may not be performing to its best. An organisation that has a culture of continual improvement presents an ideal environment for implementation of such a strategic framework.*

5.2 Identification of key processes

The process view of the organisation (as collected in preparatory phase) is to be analysed and detailed processes need to be identified and analysed as a first step. The second step is to produce an ideal process map that must exist for the organisation. This can be developed based on information on better processes deployed within other groups of the same organisation or other better performing organisations. Where an organisation considers its process model matured enough, such that it cannot identify any better process model/ its elements, it can rely on its own internal process model.

The processes that form the SAMF of an organisation are complex but interlinked. In order to ensure overall effectiveness within the organisation, each process in itself needs to be matured, well understood and applied. At the same time, it is the interplay between these processes that provide the overall strength to the organisation and improves its effectiveness and efficiencies. For e.g. even if each part of the body were performing at their best, if they are not synchronised to each other, the end result will not be as productive as it could / should be.

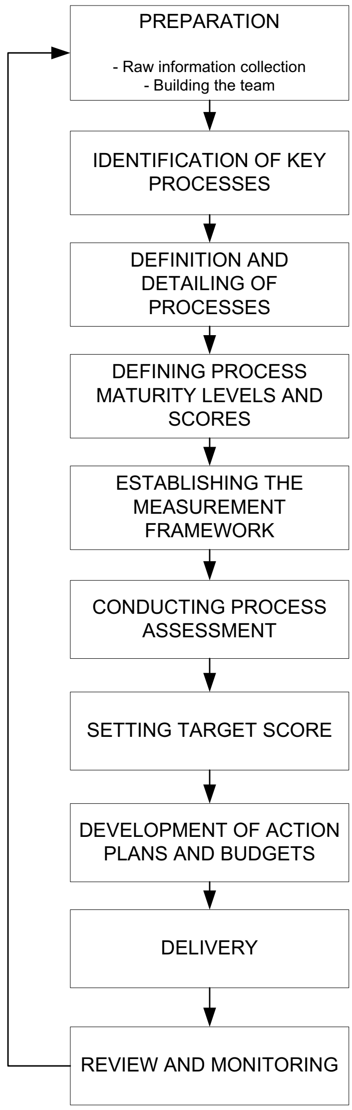


Figure 2 - Steps for Development and Implementation of SAMF

Once the information on the individual processes have been received, an overall process chart should be formed to show how individual processes link to each other to get the final outcome. This process chart needs to display all the inputs and outputs including the manner in which these interact.

A typical process model for an asset intensive organisation is shown in the figure 3. Processes must be identified for all aspects of the business from inputs down to the delivery/ outputs.

5.3 Definition and detailing of processes

The processes being identified in the last step need to be defined and detailed. This will result in a –

- Definition of each process.
- Dissection of the processes into individual process elements.
- Detailed mapping of the input – output cycle for individual processes and process elements.
- Detailing what the desired outcomes of each process will be and how will they be monitored. This must include setting appropriate performance targets that can be utilised to measure the performance of individual process.

- Process / process element owners must be identified for each process. Ideally these must be the lowest level staff responsible for the overall process within the organisation. This will ensure higher degree of acceptance.

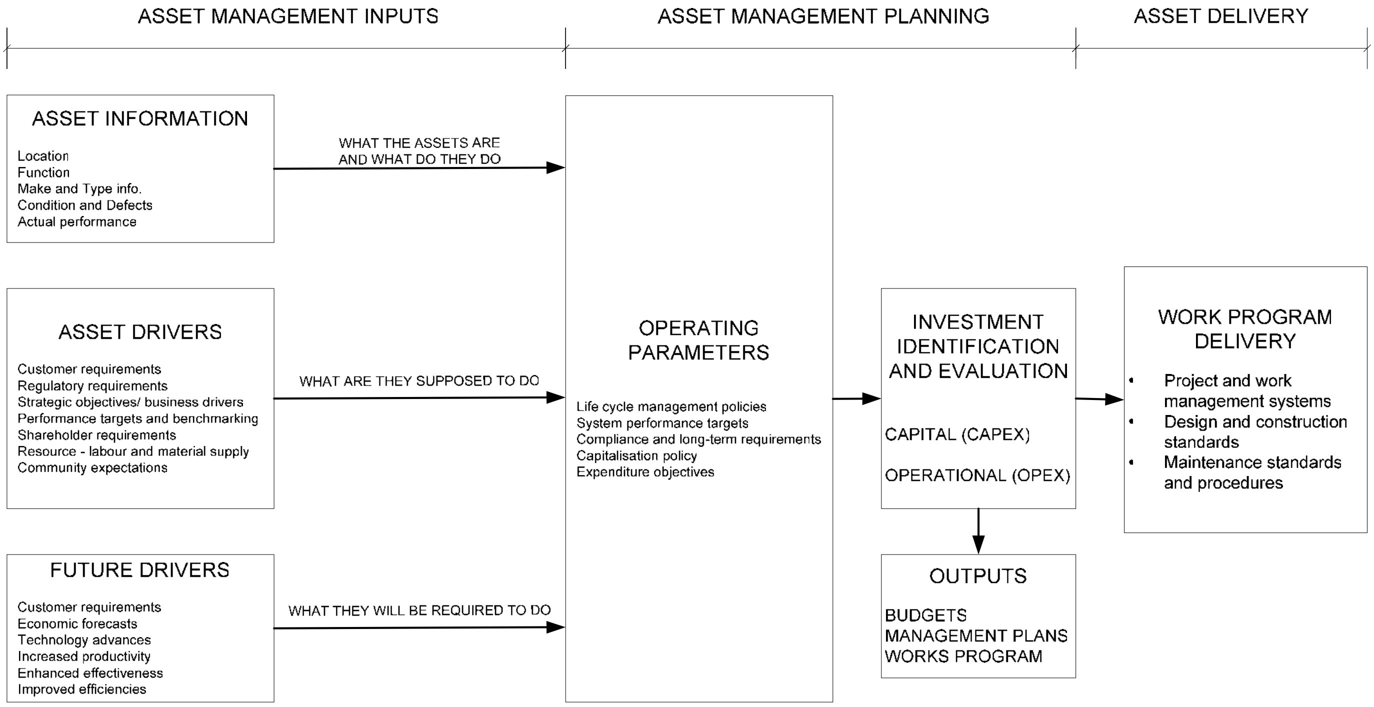


Figure 3 - Typical asset process model for an asset intensive organisation

An example of the elements to individual process is –

Process Asset Information

Process element Asset register, Asset attributes, Asset data, Systems integration

The process of splitting of the processes to elements provides a more robust platform to develop the asset management framework and subsequent delivery. The processes should be split to as many manageable process elements as possible.

5.4 Defining maturity levels and scores

This step involves defining the terminology that will be used to assess the maturity level of individual processes / elements and the related scores that will be used to mark the maturity of individual processes. A suggested terminology and score range that can be used for measurement are discussed below. These are represented in the figure 4. The terminology and scoring methodology include this measurement concept is discussed in [3] (section 2.4.3, Current Asset Management Status and Gap Analysis).

Unaware A rating that shows that the organisation and its staff are not aware of a particular process or a process element.

Score 0 - 10

Aware A rating that shows that the organisation and its staff have a reasonable awareness of the process, but do not practice (apply) it widely.

Score 11 - 30

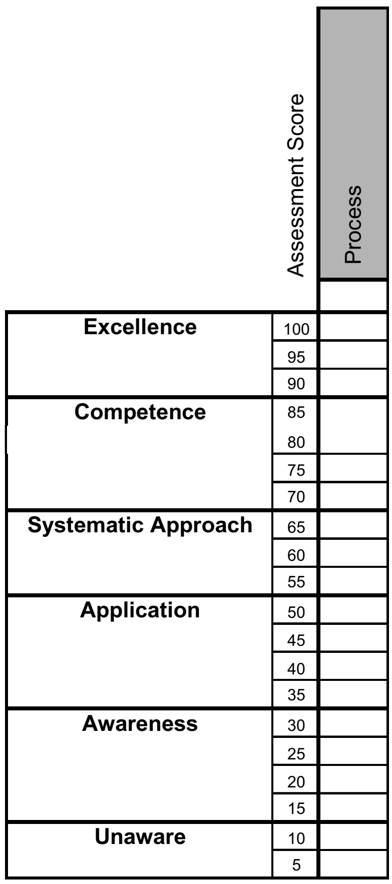


Figure 4 – Maturity Level

Application	A rating that shows that awareness of the process exists and is being applied in some cases. **Score 31 - 50**
Systematic Approach	A rating that shows that the organisations awareness of the process is very high and it is being applied consistently in almost all the cases. **Score 51 - 70**
Competence	A rating that shows that the organisation is well aware of the process, applies it systematically throughout the organisation (high compliance) and its process is matured to an extent that it can be classified as one of the best processes (when benchmarked against other similar organisations). **Score 71 - 85**
Excellence	A rating that shows that the process is applied systematically within the organisation and the process that is followed can be rated as the best in the industry. **Score 86 – 100**

5.5 Setting the measurement framework

This step involves formalising the measurement framework on which the existing maturity levels of individual processes will be gauged. The measurement framework prescribes how an individual process could be scored – for e.g. an asset register process may be considered as "unaware" if less than 80% relevant staff of organisation understand the concept, the organisation does not have a formal asset register or associated policy. At the same time, this process could be provided "systematic approach" rating if there is a formal asset register, formal asset data policy; more than 90% relevant staff understand the concept and complies with the same. This measurement framework will also assist in preparation of action plans to improve organisational process performance to improved levels.

This is a very critical step and one on which the success of the SAMF resides. This step involves establishing the benchmarks that will be utilised to score the existing maturity levels of the various processes and individual elements. This is a subjective exercise and can be very time consuming. The objective of SAMF must not be lost in this process – the objective being 'creating an environment of continual improvement and business excellence'.

The measurement framework could be set up using:

- 'absolute benchmarking' technique – that involves comparing one process of the organisation with another process within the same group or other groups within the same organisation; or
- 'relative benchmarking' technique – that involves comparing the processes of the organisation with those of another organisation with a similar or comparable operating environment.

The overall measurement framework could be a mixture of the above two techniques. The measurement framework should be formally established and endorsed by the key management staff. This will avoid conflicts and disagreements to the outcomes of the SAMF.

5.6 Conducting process assessment

This step involves formal assessment of maturity levels of each process. This requires information to be gathered on individual processes including – available formal documentation on processes and series of interviews with relevant staff (responsible for implementation of these processes). As many information sources as possible should be consulted to conduct an informed process assessment. This is a vital step as it sets the foundation to the robust framework. An inappropriate or loose assessment can lead to an inadequate framework.

The process assessment is conducted in two steps:

- The first step involves maturity assessment of individual elements of each process.
- The second step involves compilation of the individual element maturity assessment to generate an overall process maturity assessment.

The results of these assessments are then compiled to form a process assessment report. A sample of how this could be undertaken is shown in figure 5.

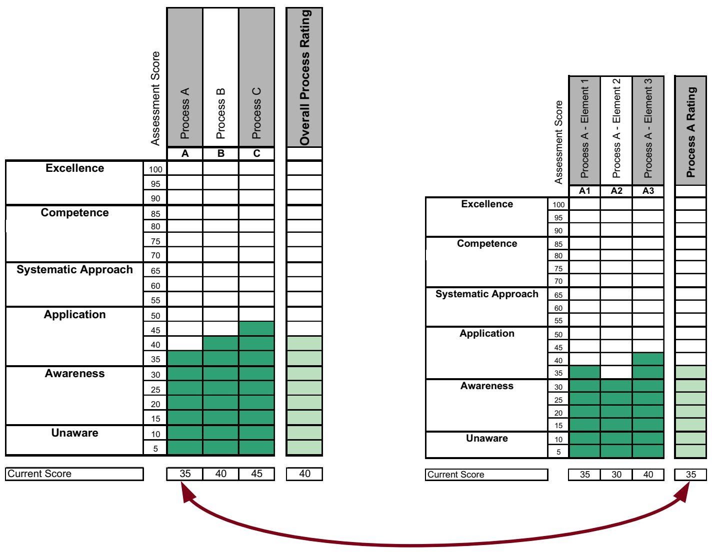


Figure 5 - Process and process element maturity assessment

5.7 Setting target score

As discussed earlier the objective of SAMF is to identify the avenues of improvement by enhancing current processes (where they exist) or developing new processes (as required) to meet the strategic objectives of the organisation.

This step in the development of SAMF involves setting the target score for each process and its process elements. Suggested time frame for implementation of the improvements is two years. The target score should be set based on an assessment of the practicality of implementing the improvements over a two-year period. The target scores are set in the context of –

- The processes that deliver the most benefit to the organisation in short-term and long-term.
- The processes that have maximum influence over the risk levels of the organisation (i.e. have more influence on reducing the risk levels).
- The processes that address the immediate business priorities of the organisation.

This step process is iterative in nature and requires recurring assessments to ensure that the target scores and associated initiatives can be delivered. Organisations need to be conscious of how much appetite they have for change. Such improvement initiatives could bring around a significant amount of change and the initiatives should be in 'bite size chunks'.

Once the target scores are set, a road map needs to be developed for incremental movements over two year period, i.e. the improvement targets for 12 months and for 24 months should be set. Figure 6 provides an illustration of the possible target score setting.

5.8 Development of action plan and associated budgets

Action plans need to be developed in order to bridge the gap between existing maturity levels of each process and the targeted maturity levels. These action plans need to be sufficiently clear and associated deliverables need to be clearly identified. The action plans should prescribe in details, the assessment criteria that will be applied to evaluate if the action has been completed or not. These criteria should be robust enough to ensure that the end result of implementation of the action is an enhanced maturity score of the individual process/ process element. The action plans should also be supported by budgets to undertake these actions. This is a substantial process improvement exercise and is adding to the value of intangible assets (process) of an organisation. The effort in this exercise should therefore be seen as an investment rather than expenditure. This will however depend on the capitalisation policy of the individual organisation.

5.9 The overall framework

The result of the above actions is a completed SAMF that enlists the organisations core business processes and the individual elements, their existing maturity levels, targeted maturity levels and the actions to be undertaken to bridge the gap. This framework is supposed to be managed as a living document and provides a key input into the overall strategic plan of an organisation.

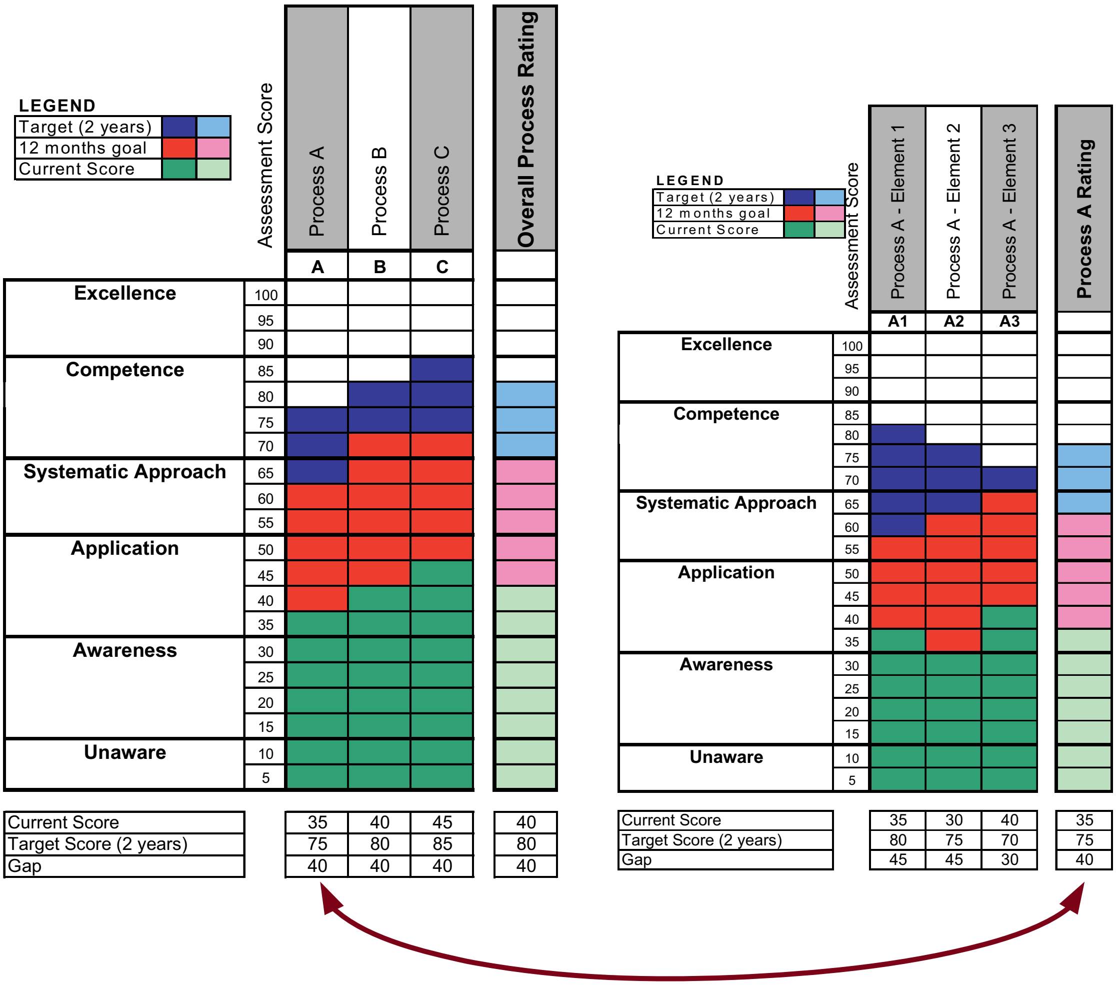


Figure 6 - Example of target score for 12 and 24 months

6 DELIVERY OF THE STRATEGIC ASSET MANAGEMENT FRAMEWORK

An open-learning environment within the organisation is a pre-requisite to the successful implementation of an initiative like SAMF. Based on the complexity of the initiatives resulting from development of the SAMF, the following steps are required to ensure their successful delivery –

- Delivery of a formal SAMF report.
- Briefing on the findings during the SAMF development to the entire organisation. This briefing should include the objectives of the SAMF, information on how the SAMF was developed, the target levels for each process (and the rationale/ objective behind this improvement exercise), expected business environment (during and after the implementation of the improvement initiatives), time frames and the overall impact on the business.
- A program management schedule should be prepared – with individual projects (improvement initiatives) clearly identified. Clear deliverables and expected quality levels of the deliverables must be prescribed.
- During the entire program, change management techniques should be planned and implemented. The change management techniques should focus not only on the implementation phase, but also on the benefits realisation phase (post implementation).
- A steering committee including a program manager and associated project team structure should be set up. Ideally, the implementation team should have a greater representation from the internal staff/ end users of the process.
- Monitoring of the program should ensure that the benefits are being realised during and after implementation.
- It should be possible to monitor the performance of the enhanced/ new processes and such information should be fed back to include in the overall framework.

6.1 Post implementation reviews

Post implementation reviews are critical for ongoing success of the SAMF. It should be remembered that, the development of SAMF is not an end of the journey. This is a never-ending journey and must involve periodic (preferred biennial) assessments of process maturity. This will ensure that best practices are continuously identified (and embraced) and any performance issues are actively addressed through feedback loop.

7 EXPERIENCES AND CONCLUSIONS

The proposed process of development of SAMF has been successfully utilised by a regulated, monopoly utility. It is acknowledged that the process is easy to use, but requires careful planning and adaptation to individual organisation's environment.

Development and implementation of a strategic asset management framework is critical for long-term success of asset intensive organisations. A process for development of the SAMF has been discussed in details, including the key aspects to be applied during implementation of this SAMF. This framework is based on the key concepts as discussed in the International Infrastructure Management Plan [3] and been practically implemented in a regulated monopoly utility.

Strategic Asset Management Framework development could result into useful insights to the business model (and its processes) and is a vital input to strategic business planning exercise. The SAMF development, implementation and continual review will assist organisations to unleash maximum value out of its resources and deliver best returns to its stakeholders.

8 REFERENCES

[1] Griffith University, Office of Facilities Management, Strategic asset management plan, web reference - http://www.gu.edu.au/ofm/publications/pdf/pub_samplan_2005-2008.pdf

[2] Department of Treasury and Finance, Treasury, Government of Western Australia, web reference - http://www.dtf.wa.gov.au/cms/tre_content.asp?ID=1269

[3] International Infrastructure Management Manual published by Association of Local Government Engineering NZ Inc (INGENIUM), National Asset Management Steering (NAMS) Group.

[4] James M. Seibert, Wade Freeman. Strategic Asset Management: an Earnings Growth Driver in Energy Delivery Business. Energy Houston Vol.4, No. 2 (2002).

Acknowledgments

The author acknowledges the guiding principles of asset management as presented in the "International Infrastructure Management Manual" that have inspired this paper. This manual presents the key concepts that have been applied by the author in the development of the Strategic Asset Management Framework (SAMF).

ISO STANDARDS FOR CONDITION MONITORING

Leith Hitchcock

Reliability Solutions, Pall Corporation QLD, Australia.

Abstract: This paper introduces delegates to the content and requirements of the new, and under development, range of ISO standards relating to condition monitoring. The paper describes the outlined processes for condition monitoring system design and implementation as well as new concepts in the fields of diagnostics and prognostics. The processes outlined also have direct FMEA links to RCM programs and can be used to implement CM programs directly linked to RCM analyses.

This paper also outlines developments in the fields of Training and Accreditation including new and upcoming standards.

Several of the new standards released and under development discussed in this paper are:

ISO 17359, Condition monitoring and diagnostics of machines – General guidelines

ISO 13373, Mechanical vibration and shock - Vibration condition monitoring of machines

ISO 13379, Data interpretation and diagnostic techniques which use information and data related to the condition of a machine.

ISO 13381, Condition monitoring and diagnostics of machines – Prognostics

ISO 18436 Part II, Accreditation of organizations and training and certification of personnel - Part II - General requirements for training and certification – Vibration Analysis

ISO 18436 Sub Parts under development for Oil Analysis and Thermal Imaging

Key Words: RCM, Condition Monitoring, Prognostics, Vibration Analysis, Thermal Imaging, Diagnostics, ISO.

1 INDUSTRY BACKGROUND

Prior to 2002 litigation based on negligence in the condition monitoring business was not largely undertaken due to the lack of standardised procedures, processes, methodologies and associated training and certification requirements and systems available in the public domain that could be used as a basis for argument. This was also compounded by a lack of knowledge within the customer base. This lack of documentation has also resulted in customer contracts with no clearly stated and/or enforceable performance criteria that could be used to form the basis for such litigation.

This situation changed in 2002 with the release of the first of many ISO Standards covering the subject of condition monitoring the most important of which are ISO 17359 Condition monitoring and diagnostics of machines – General guidelines and ISO 18436 Condition monitoring and diagnostics of machines – Accreditation of organisations and training and certification of personnel.

These two documents, and their normative references, outline the processes by which condition monitoring should be performed and also the requirements for training and certification of personnel both of which form a basis for future litigation for negligence given that these documents are now in the public domain as International Standards.

Customers using these documents for the basis of their contracts will also now be able to specify compliance to the ISO processes as mandatory as well as specifying certification standards for personnel supplied. Customers are also becoming highly educated in condition monitoring particularly with the global expansion of such maintenance philosophies as RCM, TPM, Six Sigma and a global drive toward reliability as a focus rather than maintenance cost optimisation hence the requirement for such standards.

2 USE OF STANDARDS

In general the standards being produced can be used in a wide variety of ways including:

- Managing the quality of a delivered service
- Tender or contract compliance conditions
- Evaluation of service providers
- Internal program management
- Risk management
- Insurance management
- Service provider performance evaluation
- Litigation
- Training and education

Given the ease of quoting a standard in a contract document there is a significant risk associated with lack of customer knowledge of condition monitoring principles and techniques. This lack of knowledge coupled with a lack of thorough understanding of the standards requirements may still result in a high risk service provider being selected based on the lowest price basis without regard to risk minimisation. In future far more attention is needed across industry toward Return on Investment (ROI) and risk management approaches to selection and evaluation of service providers to ensure that the service is fit for purpose, with regard to risk of failure, and has the greatest return on investment potential with respect to problem solving rather than just defect detection.

3 GENERAL PROCESS MAP

A general conceptual approach to the overall condition monitoring process is given below in Figure 1 (ISO/CD 13379). An overview of the guideline process from ISO 17359 is shown below in Figure 2.

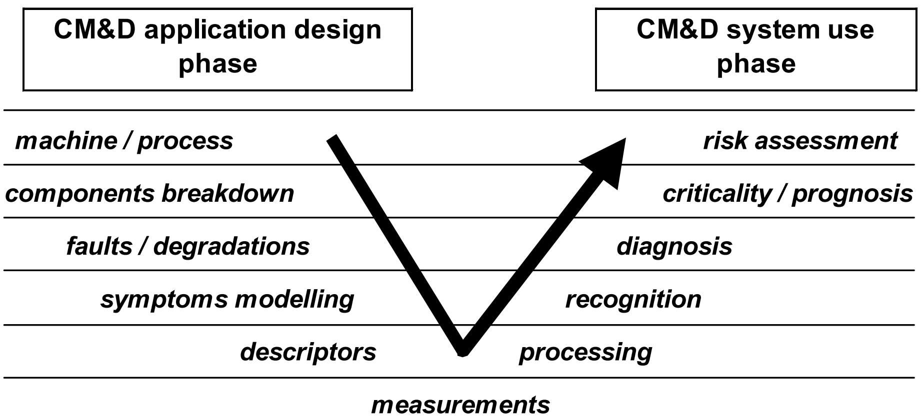


Fig. 1. Conceptual Approach.

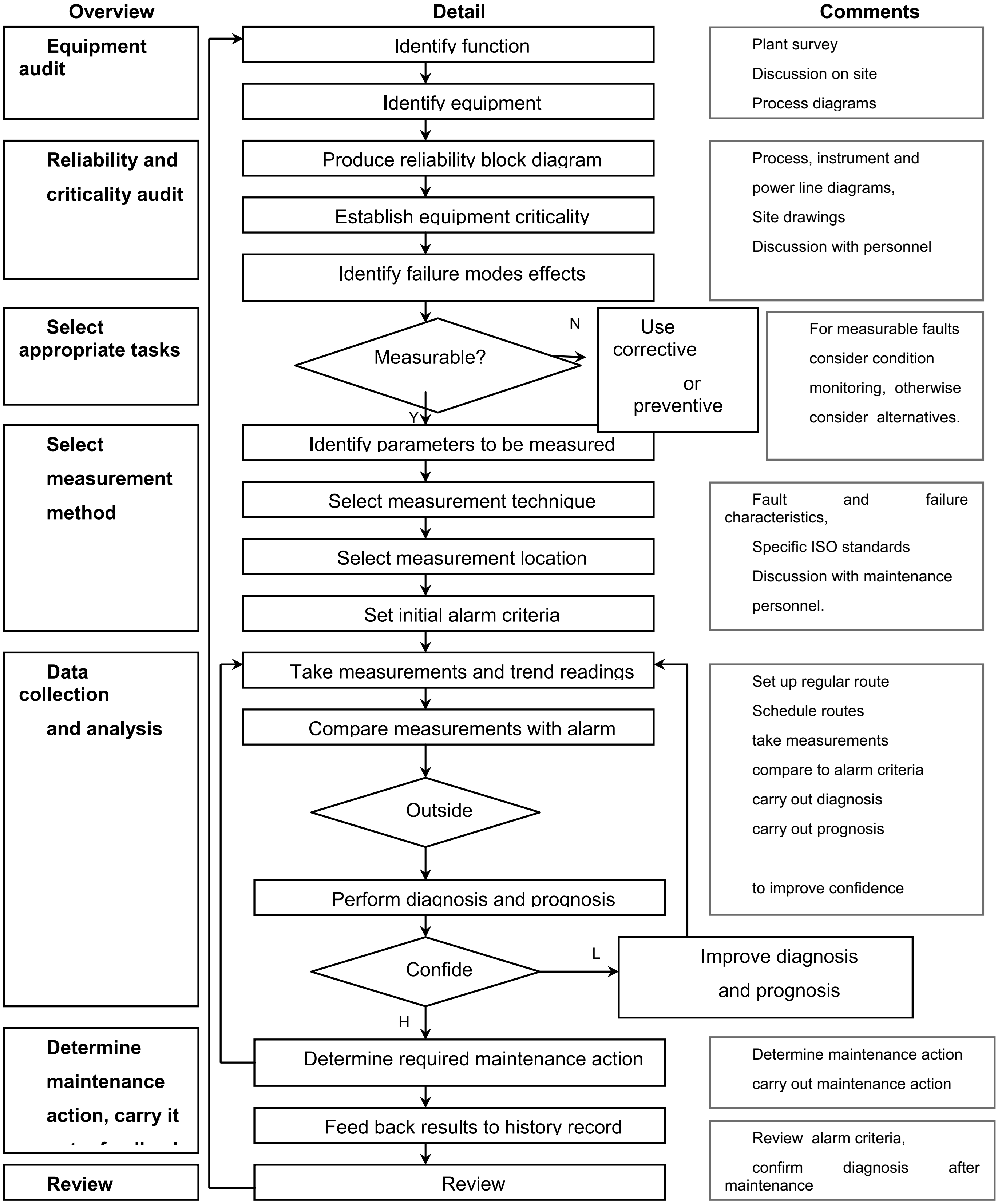


Fig. 2. ISO 17359 General approach diagram extract.

4 SELECTING THE RIGHT CONDITION MONITORING TECHNIQUE (FAILURE MODE SYMPTOMS ANALYSIS (FMSA)) – ISO 13379

4.1 FMSA process

The aim of this process is to select monitoring technologies and strategies that maximise the confidence level in the diagnosis and prognosis of any given failure mode. This methodology is designed to assist with the selection of monitoring techniques that will provide the greatest sensitivity to detection and rate of change of a given symptom. Where the confidence in a technique's sensitivity and resulting diagnosis/prognosis accuracy is questionable then the use of additional techniques for further correlation should be recommended.

This process is essentially a modification of a FMECA process with a focus on the symptoms produced by each identified failure mode and the subsequent selection of the most appropriate detection and monitoring techniques and strategies. This tool should be used in conjunction with an existing FMECA/RCM analysis that has already identified and ranked possible failure modes.

The essential elements of the FMSA process are:

— listing the components involved

— listing the possible failure modes for each component

— listing the effects of each failure mode

— listing the causes of each failure mode

— listing the symptoms produced by each failure mode

— ranking each symptom/monitoring technique combination by a detection rating, severity rating, diagnosis confidence rating and prognosis confidence rating resulting in an overall Monitoring Priority Number (MPN) rating.

— listing the most appropriate monitoring technique

— listing the estimated frequency of monitoring

— listing the most appropriate correlation techniques

— listing the frequency of monitoring for the correlation techniques

When considering monitoring strategies, the following form can also be used:

- A "Failure Mode" produces "Symptoms", which are best detectable by a "Primary Monitoring Technique" resulting in a high diagnosis and prognosis confidence when monitored at a given "Monitoring Frequency".

- Increased diagnosis and prognosis confidence can be gained by using "Correlation Techniques" when monitored at a given "Monitoring Frequency".

5 DIAGNOSIS AND PROGNOSIS PRINCIPLES - ISO 13379 AND ISO/CD 13381

Diagnosis process is generally triggered by anomaly detection. This detection is carried out by making comparison between the present descriptors of a machine, and reference values (generally called baseline values or data) chosen from experience, from the specifications of the manufacturer, from commissioning tests or computed from statistical data (e.g. long term average).

If confidence in the diagnosis and/or prognosis is low, then further verification may be required. If the confidence is high it may be possible to initiate maintenance or corrective action immediately.

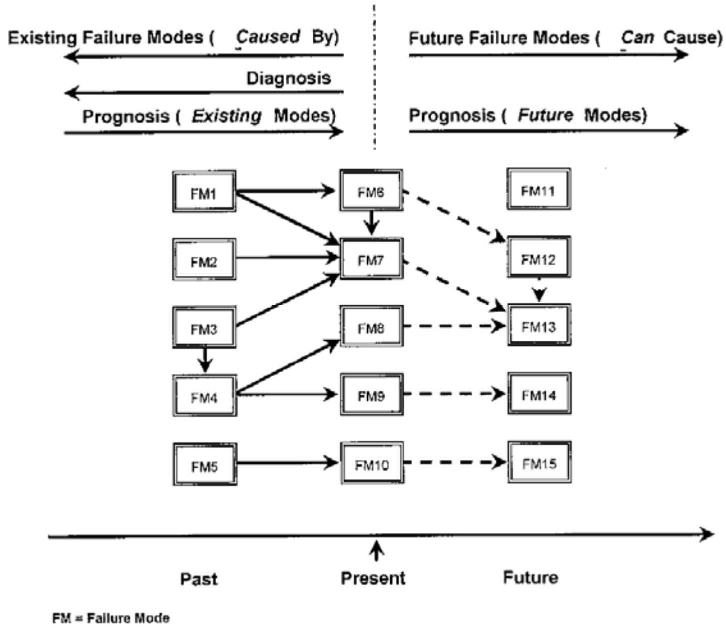


Fig. 3. Causal Tree Relationships for Diagnosis and Prognosis

5.1 Diagnosis and prognosis influence factors

Influence factors are parameters that effect the deterioration rate of a failure mode such as temperature, viscosity, clearance, load, speed, etc. Each influence factor can be considered a symptom of an existing failure mode and in Figure 3 is represented by the solid lines that connect existing failure mode trends. Influence factors also have effects on the progression and initiation of other either existing or future faults (Figure 3).

5.2 Improving diagnosis and/or prognosis confidence

In order to increase the confidence in the diagnosis/prognosis it may be necessary to carry out an iterative verification process as outlined in Figure 4. This process requires additional analysis of data using either the same technology or a different one or both in order to gain further diagnostic or prognostic information.

5.3 Setting alert, alarm and trip (shutdown) limits

The failure definition set point for a parameter/descriptor is the final value that it reaches at the point in time when the item fails. This value is normally determined historically from failure history.

The trip set point, however, is the parameter/descriptor value at which the machine is shut down and is normally less than its failure set point. This value is normally determined from standards, manufacturers' guidelines and experience. This is the value normally used to define the failed condition, however, this value is not normally reflective of the fully failed condition due to its lower set point required to prevent consequential damage or catastrophic failure. Alert and alarm limits are normally set at a value less than the trip set point.

For vibration condition monitoring, information on alarm criteria is contained in ISO/DIS 13373, ISO 10816 and ISO 7919.

Usually this value is determined based on the maintenance lead time required, however, such alert values should be cognisant of confidence level of prognosis, future production requirements, spare parts delivery lead times, maintenance planning lead time required, scope of work required to rectify faults, and trend extrapolation and projection.

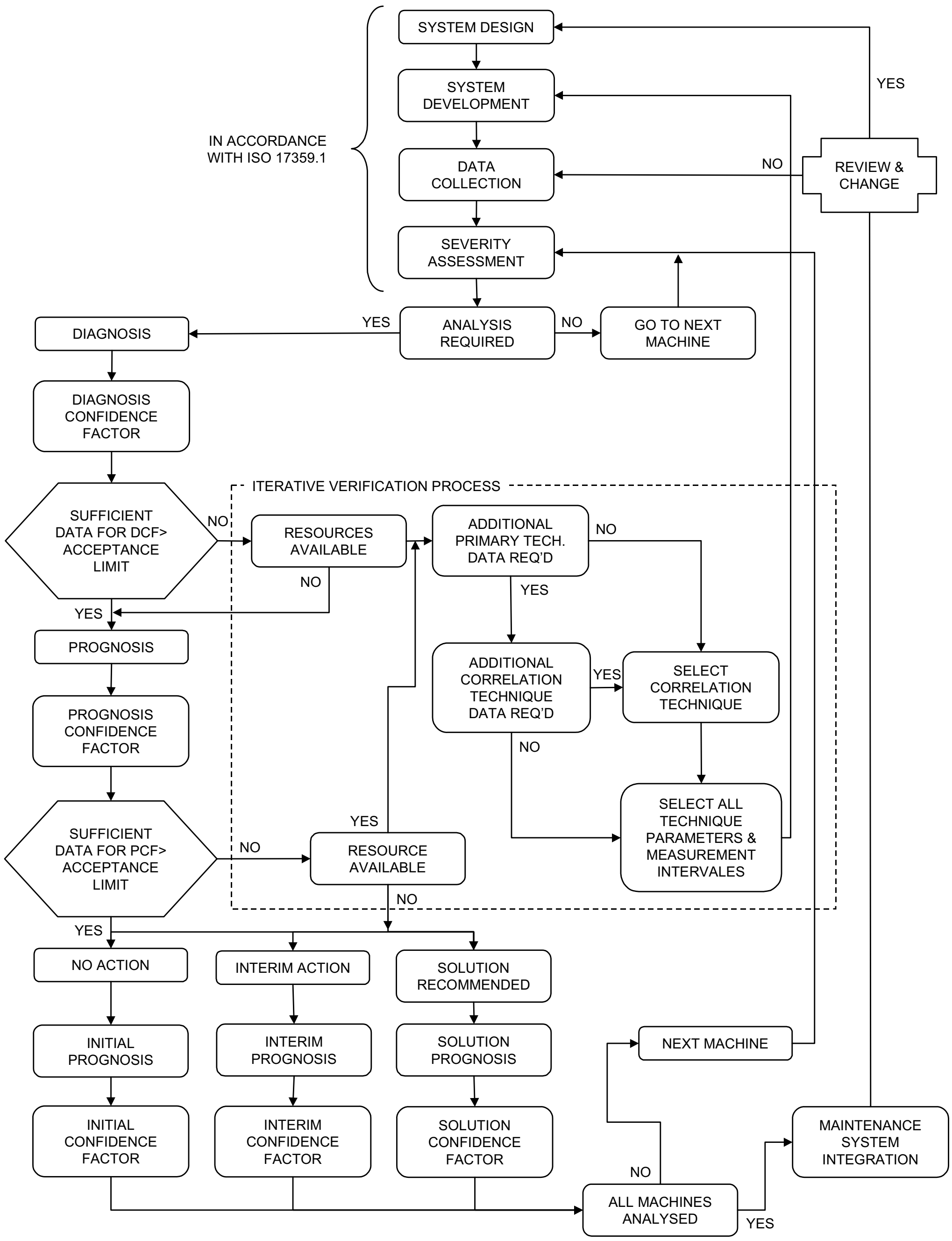


Fig. 4. Expanded Flow Chart for Diagnosis and Prognosis Confidence Level

6 TRAINING AND CERTIFICATION

The second part to running a *low risk* CM program is having qualified, experienced and certified practitioners suited to the level of risk involved in the program. The development of such standards has a global impact and is somewhat controversial given some existing international certification programs and bodies, however, TC108 SC5 has gained full liaison with BSI, BINDT, SAE, ASNT and ICML to ensure that such certification requirements are as aligned as possible to the global requirement.

To this end there are a number of ISO Standards on Training and Accreditation published as follows:

- ISO 18436-1 Condition Monitoring and Diagnostics of Machines – Requirements for Training and Certification of Personnel – Part 1 Requirements for Certifying Bodies and the Certification Process. This is a standard covering the general requirements for the certification process.

- ISO 18436-2 Condition Monitoring and Diagnostics of Machines – Requirements for Training and Certification of Personnel – Part 2 Vibration Analysis. This is a standard outlining the knowledge, experience, and training and examination requirements for a four level certification system for vibration analysts.

There are also several standards under development as follows:

- ISO 18436-3 Condition Monitoring and Diagnostics of Machines – Requirements for Training and Certification of Personnel – Part 3 Requirements for Training Bodies and the Training Process

- ISO 18436-4 Condition Monitoring and Diagnostics of Machines – Requirements for Training and Certification of Personnel – Part 4 Field Technician/Analyst. This is a standard outlining the knowledge, experience, training and examination requirement for a certification system for field based analysts.

- ISO 18436-5 Condition Monitoring and Diagnostics of Machines – Requirements for Training and Certification of Personnel – Part 5 Lubricant Laboratory Technician/Analyst. This is a standard outlining the knowledge, experience, training and examination requirement for a certification system for lubricant laboratory based analysts.

- ISO 18436-8 Condition Monitoring and Diagnostics of Machines – Requirements for Training and Certification of Personnel – Part 8 Thermography. This is a standard outlining the knowledge, experience, and training and examination requirements for a three level certification system for thermographers.

It is important to note that accreditation of Certifying Bodies is for compliance to ISO 18436 Part 1 and does not accredit the certifying organization to the body of knowledge parts. Similarly accreditation of Training Bodies is for compliance to ISO 18436 Part 3 and does not accredit the training material or delivery to the body of knowledge parts.

It also should be understood that the quality of any certification scheme in terms of validity, quality, industrial relevance and integrity is a function of its exam database and the management of such. Customers should be extremely careful with the selection of certification providers with regard to the robustness of their exam question development and management processes and as such their schemes should be thoroughly reviewed before adoption.

7 CONDITION MONITORING PROGRAM REVIEW

The CM process is an on-going process, and techniques which may not have been available, or at the time considered too costly, or too complicated, or unfeasible in some other way, may on review become feasible. Similarly the effectiveness of techniques currently being undertaken in the program should be assessed, and any techniques considered no longer necessary removed.

Similarly, warning and alarm values may need revision, due to changes in the machine such as progressive wear, ageing, modification, operation or duty cycle changes. Measured values and baselines may also change due to maintenance work, including component change, adjustment or duty change.

In certain cases the baseline may need to re-established following such changes. When maintenance actions have been carried out it is also useful to inspect components to confirm that the initial diagnosis or prognosis was correct.

The ISO Standards outlined form a basis for the design, implementation and management of a low risk condition monitoring program and as such are an invaluable resource to both suppliers and customers of condition monitoring services.

8 RECOMMENDATIONS

It is highly recommended that asset owners, service providers, and asset insurers become fully conversant with the knowledge and requirements outlined in the above standards in order to minimise the business risk associated with their activities.

Asset owners who outsource their condition monitoring should use such standards as the basis for evaluating, selecting and managing service providers and contracts. Asset owners with internal programs should use such standards to audit, validate and manage internal programs to ensure they are efficient, effective and are minimum risk.

Service providers should use such standards to design, refine and manage their programs and minimise their business risk.

Insurers should use such standards to audit and validate internal and external programs at customer facilities to ensure premium discounts or rebates and claim payouts are warranted.

9 SUMMARY CONCLUSIONS

It is foreseen that in the future these standards will form the basis for contracts, litigation and prosecution. It is not unrealistic to expect a future contract for condition monitoring services to read :

"The condition monitoring program with be designed in accordance with the requirements of ISO 17359, and where applicable ISO 13373 with the program designed by a person certified to a *minimum* of the second highest level for each technology category applied.

The selection of appropriate techniques shall be in accordance with ISO 13379. The vibration analysis component of the program shall be compliant with ISO 13373, ISO 10816, ISO 10817, ISO 7919, ISO 5348, and any other ISO standards relevant to the technique and application. The thermal imaging component of the program shall be compliant with ISO 18434. The lubricant analysis component of the program shall be compliant with ISO 14830. The performance monitoring based component shall be compliant with ISO 13380.

Diagnostic processes shall be compliant with ISO 13379 and ISO 13373. Prognostics and all reporting requirements shall be in accordance with ISO 13381.

Personnel shall be trained and certified to ISO 18436 with the program managed by a person certified to a *minimum* of the second highest level for each technology category applied.

The program and its compliance to standards will be audited and validated using the above stated standards. The stated standards, and their normative references, shall form the basis of any compliance-based litigation and/or prosecution."

This will become the norm particularly when such certification programs become fully mature and if the consequences of a missed failure include injury or death.

It is not also unrealistic to see a future where one of the key business drivers for standards implementation and compliance is the insurance sector.

As a minimum it should be recognised that standards represent the condensed and refined knowledge of a large group of international experts and as such contain the distilled knowledge of global experts. If for no other reason than education and learning, and a desire to improve, then such standards are a wealth of knowledge for practitioners.

By their very nature the use of such standards can result in the standardised implementation of a highly effective, highly efficient, low risk, optimised condition monitoring program that consistently produces repeatable results in a cost effective manner. Surely this alone is sufficient incentive to use them!

INTEGRATING ROOT CAUSE ANALYSIS METHODOLOGIES

Leith Hitchcock

CPEng CMRP, PALL Corporation

Abstract: Many Root Cause Analysis (RCA) methodologies have specific applications and limitations and in some case for complex machinery investigations they can be combined and enhanced for better results. Typical methodologies that can be combined effectively are Kepner Tregoe, Causal Tree Analysis (Apollo), Fault Tree Analysis, Logic Tree Analysis, Barrier Analysis, and Human Performance Evaluation amongst others.

The difficulty with many RCA methodologies currently available today is that by themselves they may not result in complete and efficient analyses and effectively implemented solutions for complex machinery problems. Most methodologies also need expert facilitation to achieve results. Many methodologies today also have widespread general application yet can be further enhanced for asset specific investigations. Furthermore the can be effectively combined for a more effective and efficient approach.

Using Life Cycle Management principles the concept the all root causes can be classified into one of five categories: Design, Manufacturing, Installation, Operation and Maintenance, can be used to enhance data gathering, problem identification, causal tree selection and management and solution design. This concept is also applicable to both logical (KT) and graphical (CT) methodologies.

Key Words: Fault Tree Analysis, RCA, Fault Tree Analysis, Barrier Analysis, Kepner Tregoe, Life Cycle Management, Solution Design, KT, CT.

1 THE PREPARATORY PHASE

The following flowchart is adapted from Reference 1 and outlined the key sequences in machinery RCA approach:

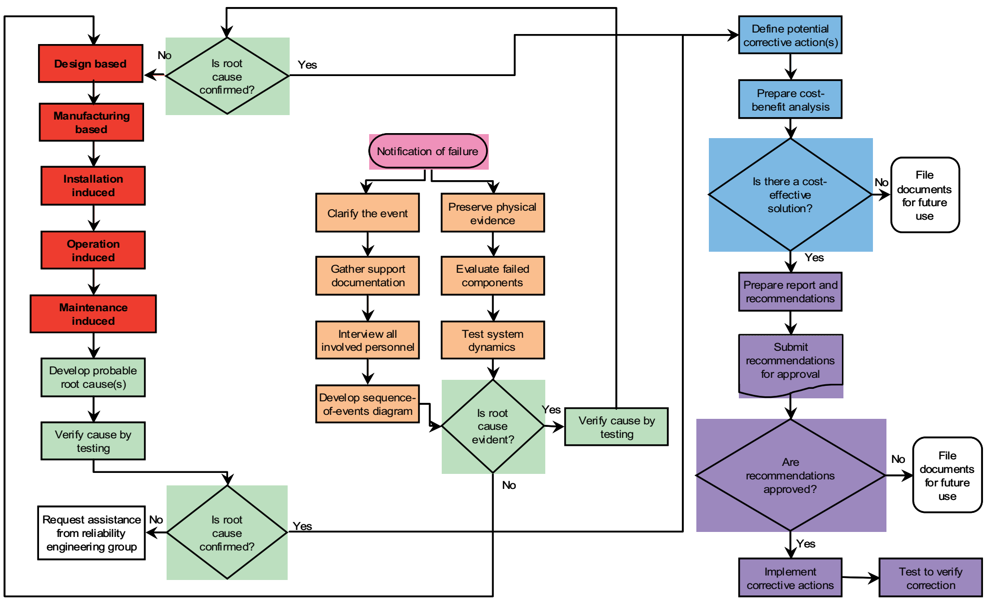


Figure 1. General Machinery Investigation Process (1)

The conceptual refinement is to drive the five basic root cause categories of Design, Manufacturing, Installation, Operation and Maintenance (DMIOM) into the RCA process via specific focussed areas of investigation.

2 THE BASIC TECHNIQUES

The following table is adapted from the US Government's DOE RCA Document[2] and outlines the applications and limitations of some for the generally used techniques:

Method	When to use	Advantages	Disadvantages	Remarks
Cause & Effect Chart Analysis *Fault Tree Analysis* *Logic Tree* *Fishbone Diagram*	Use for multi-faceted problems with long or complex causal factor chain. Fault or Logic Tree can be used when the problem involves logic or control issues.	Provides visual display of analysis process. Identifies probable contributors to the condition. Fault and Logic Tree capable of analysing logic issues with control systems or processes	Time consuming & requires familiarity with process to be effective. Fault and Logic Tree's can be extremely complex	Requires a broad perspective of the event to identify unrelated problems. Helps to identify where deviations occurred from acceptable methods. May need specialists in control systems and logic for Fault or Logic Tree methods.
Change Analysis	Use when cause is obscure. Especially useful in evaluating equipment failures.	Simple 4 step process.	Limited value because of the danger of accepting wrong, "obvious" answer.	A singular problem technique that can be used in support of a larger investigation. All Root Causes may not be identified.
Barrier Analysis	Use to identify barrier & equipment failures & procedural or administrative problems.	Provides systematic approach.	Requires familiarity with process to be effective.	This process is based on the MORT Hazard / Target concept.
Management Oversight & Risk Tree Analysis (MORT)	Use when there is a shortage of experts to ask the right questions & whenever the problem is a recurring one. Helpful in solving programmatic problems.	Can be used with limited prior training. Provides a list of questions for specific control & Management factors.	May only identify areas of cause, not specific causes.	If this process fails to identify problem areas, seek additional help or use Cause & Effect analysis.
Human Performance Evaluation	Use whenever people have been identified as being involved in the problem cause.	Thorough analysis. Looks at systemic and human aspects to failure.	None if the process is closely followed.	Requires training.
Kepner-Tregoe Problem Solving & Decision Making	Use for major concerns where all aspects need thorough analysis. Can be used as a general framework.	Highly structured approach, focuses on all aspects of the occurrence & problem resolution. Disciplined solution development process.	More comprehensive than may be needed.	Requires training.

Table 2. Method Applicability Table [2]

3 USING KEPNER-TREGOE AS A FRAMEWORK

Although the Kepner-Tregoe (KT) [3] methodology is comprehensive this degree of detail is often required for complex or extensive machinery failure investigations and warranty claim work. Using KT as a framework allows facilitators to better acquire and prepare information prior to carrying out such methodologies as Causal Tree[3] or Fault Tree[4]. This is important when the sequence of events, failure mechanism relationships and failure mode interrelationships must be understood in detail. Proper data acquisition prior to commencing the RCA also improves the processes effectiveness and efficiency.

A diagrammatic process is detailed below in Figure 2 detailing how KT can be used as a framework with other graphical techniques inserted for more detail failure mechanism/logic analysis.

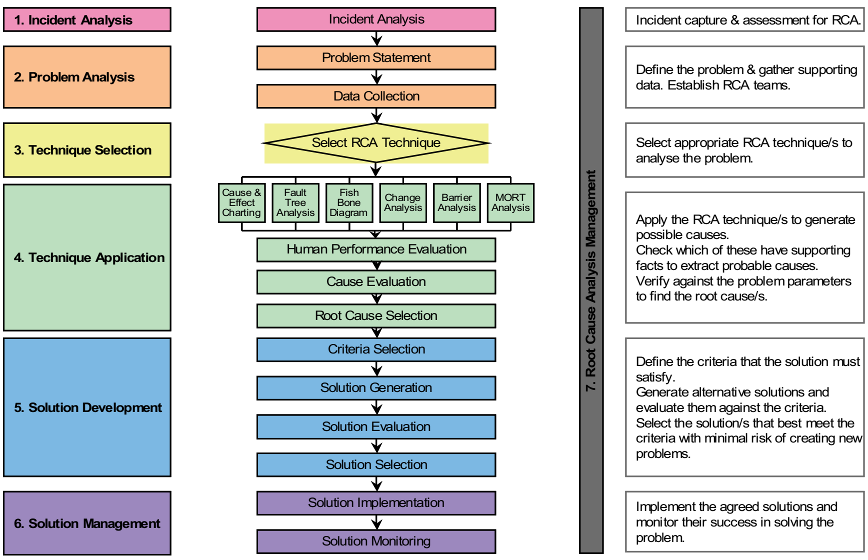


Fig. 3. A diagrammatic approach to using KT as a framework for RCA methodology integration.

The KT methodology, however, can be further improved by driving the DMIOM concept through the data acquisition, problem analysis, and solution design areas of RCA. One way to do this is to carry out the question challenge for each DMIOM category. An additional improvement is to expand the data acquisition/problem analysis phase with the basic elements from the change analysis methodology to give an additional "deviation" and "effect/consequence" dimension to the analysis as can be seen below in Table 2.

For each DMIOM category	*IS*	*IS NOT*	*Deviation*	*Effect/ Consequence*
Identity (WHAT)				
Location (WHERE)				
Timing (WHEN)				
Extent (SIGNIFICANCE)				

Answer the following questions and fill in the chart:

IS…

Identity - What item specifically has trouble? What is wrong with it?
Location - Where on the item did it happen? Where was the item located?
Timing - When did it happen - time, before / after, point in cycle?
Extent - When it happens how much is affected? Any pattern?

IS NOT…

Identity - Are there similar items? How are they affected?
Location - What parts are unaffected? Are others having trouble?
Timing - What are other likely times? Is it happening then too?
Extent - Is some portion consistently not involved? Is this usual?

Table 2. The expanded KT methodology with DMIOM category and change analysis embedded.

4 HUMAN ERROR

Human Performance Evaluation has a key role in linking solutions to DMIOM categories where there is human involvement. In the case of machinery failure human error is prevalent and in order to make solutions permanent a focus on the systemic and human dimensions needs to be carried out.

A systematic view of human error is outlined in Figure 4, which details the key areas for investigation into human and systemic error.

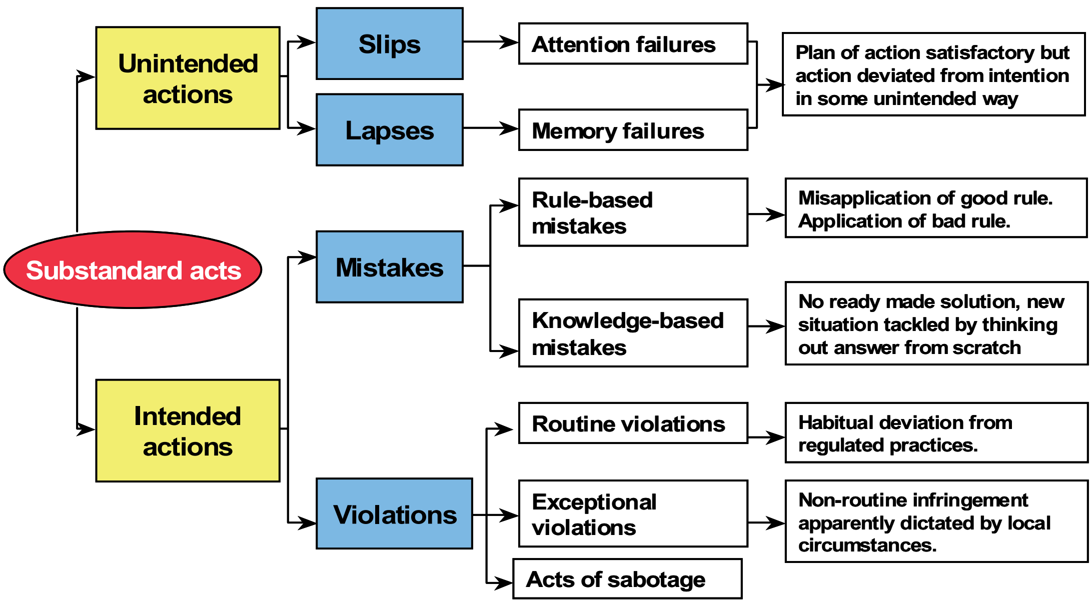


Figure 4. Human Error Sources

The analysis of human error as part of an asset RCA provides a necessary link to Incident/Accident investigation techniques used in health, safety, and environmental incident investigations where such techniques as Barrier Analysis [(2&5)] and MORT [(2&6)] are used.

5 MODIFYING CAUSAL TREE METHODS

Causal Tree analysis can often become difficult if a complex problem is not initially broken down into separate areas of analysis. This in itself poses a problem from a facilitation viewpoint in that sometimes the process can end up with several teams investigating the same fault tree. The classic example of this from a machinery viewpoint is "lubrication failure".

The author first used a matrix system in 1990 to regain control over a complex turbine investigation with great success. The system basically splits a complex investigation into several logical areas of analysis and each area is given a sequential number. The causal tree sheets (when using the Apollo method (7) for example) are then numbered and each fault tree is laid over an alphanumeric matrix such that each individual failure mode (Post-it note) can be identified using sheet number and Cartesian coordinate.

When simultaneous analyses are carried out the facilitator ensures that when a common cause is found it is only analysed once from either one fault tree or as a separate tree. Whenever this tree is referenced the initiating cause is marked with "TO Sheet X, Row Y, Column Z)" and the analysis tree is referenced "FROM Sheet A, Row B, Column C". The reason for this becomes clear when as the analysis progresses key failure modes become referenced multiple times. A failure mode that receives multiple references is either a Root Cause or a Significant Contributing Cause and must be dealt with by the solution. Quite often these multiple reference failure modes become the focus for quick fixes or interim solutions as they themselves influence multiple other failure mechanisms, as is the case for "lubrication failure" in machines investigations.

Further modifications can be made by incorporating "and, or, if" logic indicators on the causal tree branches, however, this can cause unnecessary complication of the process. This process is useful particularly in the case of conditional "if" review where an event wont take place until a preceding condition has been met.

6 CONCLUSION

Like any maintenance activity the different root cause analysis methodologies should be viewed as tools in a toolbox and as such any RCA program should not rely on just one method in isolation but rather have expertise in several key methodologies that can be deployed to suit the type of RCA required.

This concept of deploying a task oriented method is the general principle behind the US Governments DOE document.

When such techniques are used for machinery root cause analysis investigations users should consider further refinements as outlined in this paper in order improve the effectiveness and efficiency of investigations. Such processes and modification should, however, be based on the criticality of the failed asset, the risk associated with its failure, and the return on investments expected from the RCA.

7 REFERENCES:

[1] Mobley, R.Keith. Root cause failure analysis (1999). Butterworth-Heinemann, Woburn MA, USA. ISBN 0-7506-7158-0.

[2] DOE-NE-STD-1004-92, DOE Guideline, Root Cause Analysis Guidance Document (1992), US Department of Energy, Office of Nuclear Energy, Washington DC, USA.

[3] C.H. Kepner and B.B. Tregoe. The New Rational Manager (1981). Princeton Research Press, Princeton, NJ, USA.

[4] BS5760: Part 7: 1991, Reliability of systems, equipment and components, Part 7. Guide to fault tree analysis.

[5] Barrier Analysis (1995). Technical Research and Analysis Centre, Scientech, Inc. Idaho Falls, ID, USA. SCIE-DOE-01-TRAC-29-95.

[6] N.W. Knox and R.W.Eicher (1983). Mort Users Manual, SSDC-4, Rev.2, System Safety Development Centre, EG&G, Idaho, ID, USA.

[7] Gano, Dean L. (1999). Apollo Root Cause Analysis. Apollonian Publications, Yakima, Washington, USA. ISBN 1-883677-01-7.

MAINTAINING KNOWLEDGE ASSETS

Karl Matthews [a] and Howard Harris [b]

[a] *School of Management, University of South Australia; Principal Consultant of CBDConsultancy.com*

[b] *School of Management, University of South Australia*

Abstract: The importance of organisational information and knowledge in developing advantage and sustaining business competitiveness is becoming more widely recognised. Various standards have recently been developed to assist businesses create, store, access and reuse their knowledge. These standards include ISO 5037 Knowledge Management, ISO 9001 Quality Management, ISO 4801 Health and Safety, ISO 14000 Environmental Management and ISO 15489 Records Management. The management of knowledge, and the requirements of the standards, has led to the development of new management systems specifically designed to assist maintain organisational knowledge assets. However, it is considered that unless these new systems are implemented effectively they will not succeed in assisting managers maintain information and knowledge assets. This paper discusses this modern issue of information and knowledge asset maintenance. It includes a case study and taxonomy of the issues that influence the implementation of new management systems. These will provide insights into practical implementation issues and how they can be overcome. Ultimately, it is through better implementation of information and knowledge management systems that managers can better maintain their knowledge assets. The objective of this paper is to refine models that assist managers maintain organisational knowledge assets.

Key Words: Knowledge, Wisdom, Asset, Maintenance, Management, Change, Implementation, Agility

1 THE EVOLUTION OF KNOWLEDGE ASSETS

Over the 800 lifetimes of human existence, only during the last 6 have we benefited from the printed word. Only in the last two has anyone used an electric motor, and the overwhelming majority of all the material goods we currently use in daily life today have been developed within the present lifetime [1]. It is considered that the evolution of human knowledge has been a significant enabler to the technology development that has occurred over the last two 'lifetimes'.

This knowledge forms a significant part of organisational asset capital, being the most powerful engine of production and thus an asset requiring management [2]. Human knowledge is considered to be an Intangible Intellectual Capital Asset Resource [3]. By simple definition, an asset is a resource controlled by an enterprise as a result of past events and from which future economic benefits are expected to flow to the enterprise [4]. Intellectual capital assets (including knowledge) are as invisible as the roots of a tree, however these roots form a proportion of the entity and significantly affect its future performance [5]. The asset of Intellectual Capital is traditionally defined as having three parts being the human aspects, intra-organisational structures and the external environment [6].

According to Ahonen (2000), intangible assets can be divided into two classes, namely generative and commercially exploitable intangibles, which enable intellectual capital based value creation. The commercially exploitable intangibles consist, for example, of production, immaterial property rights, customer capital, demand and management. These commercially exploitable intangibles can be acquired or self generated through generative intangibles such as human capital, internal structures and external structures. These generative intangibles describe the capacity of a firm to produce commercially exploitable intangibles [7]. Knowledge assets therefore exist in both classes.

There are two different kinds of knowledge, that is, tacit and explicit [9]. Tacit knowledge resides in a person's mind and may include aspects of culture, or 'ways of doing things' [10]. It is personal and unarticulated, therefore, difficult to describe, communicate and share with others. It includes lessons learned, know-how, judgment, rules of thumb, and intuition [11]. Explicit knowledge is recorded as information in a document, image, film clip or some other medium [10]. It is clearly formulated or defined, easily expressed without ambiguity or vagueness, and codified and stored in information systems [11].

In Nonaka and Takeuchi (1995) the interaction between tacit and explicit knowledge has been expressed by means of the SECI model, which consists of four different modes of knowledge conversion: socialisation, externalisation, combination and internalization. Knowledge creation is a spiralling process in which different modes of knowledge conversion follow on each other. This process is called the epistemological level of knowledge creation. The SECI process is shown in Figure 1. This epistemological spiral of the SECI process expands both horizontally and vertically across the organisation, into new ontological levels. These two dimensions constitute knowledge creation [8].

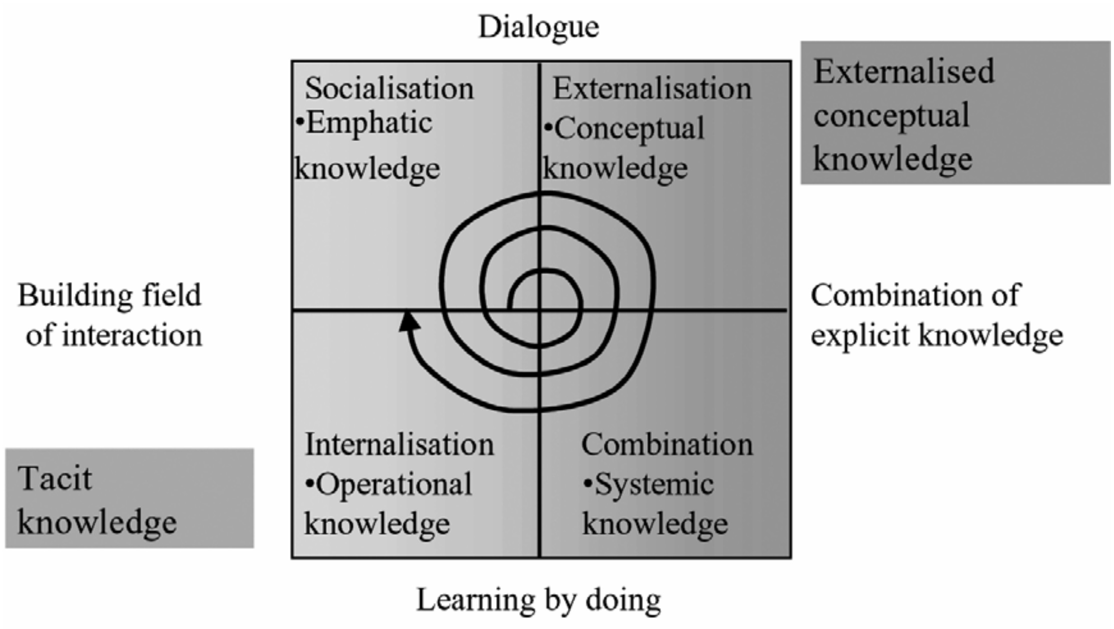


Figure 1 : The spiral of knowledge in the SECI model [8]

Thus the process of knowledge management begins with this knowledge creation and progresses to knowledge capture and storage, knowledge refinement, knowledge distribution, knowledge use, and monitoring of the entire process, which should then impact the beginning of the process [12]. In 1996, one view was that knowledge management involved knowledge creation, followed by knowledge interpretation, knowledge dissemination and use, and knowledge retention and refinement [13]. The American Productivity and Quality Center (APQC) [14] defined knowledge management as "the strategies and processes of identifying, capturing and leveraging knowledge" to enhance competitiveness. Miller considered that the knowledge management process is concerned with capturing an organisation's know-how and know-what through creation, collection, storage, distribution, and application [15]. Svieby [16] modelled the Create, Store Use, Reuse, Dispose cycle. In 2003, the Australian Standard in Knowledge Management clause 1.4.1 determined that the knowledge management process consisted of four activities: creating, discovering and acquiring knowledge; capturing and storing knowledge; presenting, distributing and sharing knowledge; and revising and disposing of knowledge [10]. Notably, clause 4.3.1 of the Australian Standard on Record Keeping details a similar process, being capture, registration, classification, classification of access, security and disposal, storage, use, and disposal [17]. Figure 2 presents a knowledge lifecycle model developed from this literature review.

.

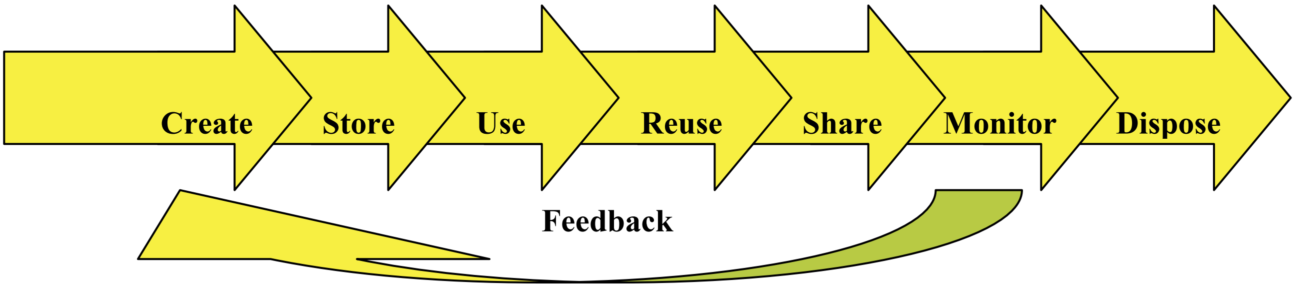


Figure 2 : Knowledge life-cycle model

1.1 Knowledge management

The key goal of knowledge managers should be to understand these knowledge management (lifecycle) processes and incorporate them into business activities. [12]. As Martin rationalises, the real point of efforts at understanding intellectual capital is not a matter of measurement or financial reporting but of management for creation of value. [18]. Knowledge management is the process of critically managing knowledge to meet existing needs, to identify and exploit existing and acquired knowledge assets and to develop new opportunities [19]. The two objectives of knowledge management are firstly to make the organisation act as intelligently as possible in order to secure its viability and overall success; and secondly for the organisation to otherwise realise the best value of its knowledge assets [20].

There are actually considered to be three major schools of thought on what knowledge management is [21]. One school suggests that knowledge management is primarily information technology enabling people to share knowledge. A second school suggests that knowledge management is more of a human resource issue with emphases on organisational culture and teamwork. A strong, positive organisational culture is critical to promoting learning, development and the sharing of skills, resources, and knowledge. The third school promotes the development of processes to measure and capture the organisation's know-how. Processes do not necessarily need to involve the use of information technology.

There has also already been three generations of knowledge management [22]. The first generation focused on defining knowledge management, investigating the potential benefits of KM for businesses, and designing specific KM projects [23]; [8]; [24]. The second generation of KM started to emerge around 1996 with many corporations setting up new jobs for KM specialists. The different sources of KM combined and expanded to touch frameworks [25], operations and practices [26] and advanced technologies [27]. Second-generation KM emphasised that KM is about systemic organisational change co-developed with management practices, measurement systems, tools and content management. Resulting from these new insights and practices, the third generation of KM emerged, according to [28] to integrate with the enterprise's philosophy, strategy, goals, practices, systems and procedures and how it becomes part of each employee's daily work-life and motivation. The third generation seems to emphasise the link between knowing and action [29].

The authors consider that in this dawn of a new millennium, a fourth generation of Knowledge Management is now evolving. This is supported by the recent 'third wave of environmentalism' as recognised by Holgrem [30]. This fourth generation seeks to integrate the synergistic philosophies of Quality and Knowledge Management, and apply wise decision making discipline to implement only the highest quality knowledge. This would support the principle of 'real quality' as proposed by Heeks [31]. Although Hawking dreams of a day beyond unified theory when we have a complete understanding of our existence and the mind of God [32], such a quest for such 'knowledge' is considered suicidal unless wisdom prevails. Multi-loop learning inevitably converges human knowledge to the wise practice of organic sustainability. The practice of fourth generation knowledge management is to enable that outcome. It is considered that this fourth generation evolution of knowledge management would therefore actually be more appropriately coined 'Wisdom Management'.

The most common objective of first adopters of those earlier generation knowledge management systems was implementing some sort of knowledge repository. The objective of this was to capture a particular type of knowledge (eg best practices, sales information, lessons learnt, competitive intelligence, plans etc) for later and broader access by others within the same organisation. [12]. Since the implementation of these early knowledge management systems, organisations have been better enabled to create, store, refine, distribute, use, monitor and feedback knowledge, as per the Figure 2 lifecycle.

However, as the human knowledge (and technology) further evolved forward, ongoing maintenance of the knowledge assets and the system itself is required to retain the asset in an optimised state to enable organisational competitiveness. Therefore, this raises a requirement for organisational knowledge condition monitoring and maintenance techniques.

2 MAINTENANCE OF KNOWLEDGE

This section considers 'familiar' maintenance principles being applied to knowledge assets. To discuss a maintenance system the object system must first be defined [33]. In this paper, the object system being maintained is 'knowledge'. With the rapid evolution of organisational knowledge, knowledge maintenance has become a challenging activity for most competitive organisations. There is growing recognition in the business community about the importance of knowledge as a critical resource for organisations [34], [35]. In this new economy, individuals and companies are obliged to focus on maintaining and enhancing their knowledge asset in order to innovate [22]. Organisational knowledge evolution (learning) is a result of an efficient and effectively designed and implemented Quality Management System [36]. Such knowledge evolution enables continuous improvement [37], and thus further knowledge evolution. As with technical processes and production capacities, the process of quality management is also liable to wear as parts of the system become ignored or go out of date [38]. Maintenance activities are necessary to "repair" the quality knowledge assets as discrepancies arise. The outcome of Knowledge Management is a continuous implementation of knowledge improvements.

The maintenance of the knowledge asset is considered by linking maintenance concepts [39], [40] with concepts of quality and knowledge management [41], [8], [42]. The maintenance of knowledge can be through either built-in or non-built-in maintenance. External and internal audits are examples of built-in maintenance. The results of 'built in' audits and inspections affect the performance of organisational knowledge development, in the sense that explicit operational process performance quality is reviewed, and changes introduced. Such built-in actions enable maintenance of the organisational knowledge. If the activity has been left to chance, performed in addition to the normal quality control activities, it is of the non-built-in type. Often, non built-in maintenance is performed to put right (or to prevent) the poor functioning of the maintenance management (control) system itself.

The maintenance system comprises both formal and informal elements. The formal elements include management systems and methods, organisational structures, information systems, and technology necessary to implement the generic maintenance tasks within an industrial enterprise. The formal management system essentially specifies when to deal with one or more of the maintenance tasks, what to do when a task is dealt with, and who should do it considering formal collective knowledge and know-how. Informal elements of a maintenance system include the actors of maintenance. These actors include the individuals performing maintenance functions (ie. technicians, operators, and managers) as well as the influence of the corporate culture. Individuals of a maintenance organisation act and decide using their knowledge, skills and know-how, their motivation and attitudes [33].

Analogous to maintenance models for technical systems and products, the notions of preventive maintenance and corrective maintenance can also be applied to the maintenance of knowledge [38]. Preventative-based maintenance includes condition based or user based maintenance. Condition based maintenance is applied if the result of an observation violates defined norms. The observations may be done on a continuous or periodical basis. Inspections, audits, and monitoring can trigger condition-based maintenance of existing organisational knowledge. User based maintenance can be applied if there exists a relation between the decrease in the condition of the system and the intensity of use. The intensity of use is somehow observed and on the basis of this observation it is decided if it is necessary to maintain the system (or its parts). This maintenance method is most applicable to the knowledge that is (at least in principle) liable to ageing. Knowledge loss (losing documents or staff) is one example, and knowledge relevance (referencing most up to date information) another. On the basis of the results of user observations, a “knowledge repair” can be implemented. Corrective, failure based maintenance can be applied when certain aspects of the quality management system do not work as they should. Maintenance takes place only after a failure occurs. For example, when a customer complaint indicates opportunity for knowledge system improvement, maintenance of the system to reduce similar complaints can be implemented. These knowledge maintenance methods are depicted in Figure 3.

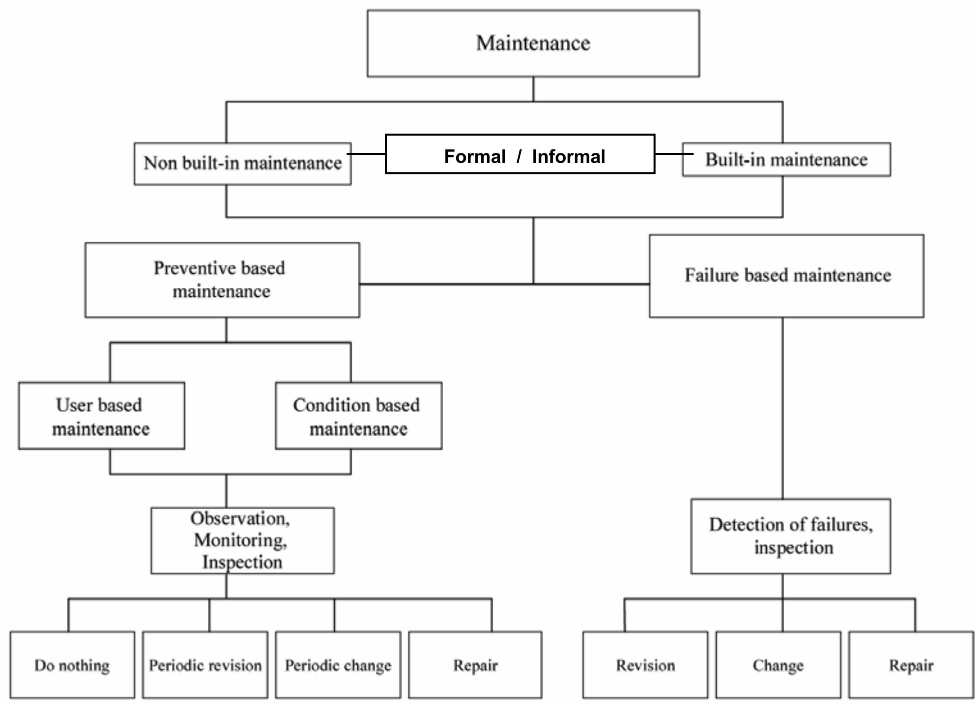


Figure 3: Knowledge Maintenance Methods

3 IMPLEMENTING TOOLS TO MAINTAIN THE KNOWLEDGE ASSET

Business involves competitive advantage [43]. Competitive advantage requires innovation [44]. Improvement and innovation play a crucial role in securing sustained competitive advantage [43]. The continual improvement of quality brings continual rise in productivity [45]. Innovation and continual improvement are a main outcome of knowledge management, which benefits competitive advantage. Therefore, it can be deduced that successful implementation of knowledge asset improvement positively influences business excellence. As organisations further appreciate these benefits of the knowledge asset on business competitiveness, engineers are kept striving to better maintain the knowledge asset.

The continuously changing and improving knowledge assets require implementing in order to deliver the competitive advantage capabilities and organisational excellence. It is vital that managers understand the issues that influence the implementation of these improvement changes, and how to offset any issues before they become dysfunctional. Successful knowledge management must ensure the balancing of the elements of People, Process and Product (Technology & Content) to fit with the organisation's capability and culture [10]. Likewise, it is considered that all three elements of people, product and process must also be considered to achieve successful implementation of a knowledge management change.

3.1 An emerging implementation model

Managers should be aware that these people, product and process issues can occur at any stage of the change implementation. Rogers recognised five stages of implementation; two stages prior to the decision to adopt the change, and three stages once that decision has been taken. The final stage occurs when everyone is using the new system and it has become routine [46]. The individual participation in the process of implementing change is also considered to follow a change curve from denial through to commitment [47]. The capability maturity model (CMM) provides a five-stage classification for describing and assessing the extent to which an organisation has become proficient in adopting a particular technology or system[48], and Satir has developed a change management philosophy [49]. These models – Rogers' stages, Gaffe's change grid, capability maturity and Satir's curve – are not in conflict but rather detail different aspects of the change implementation process. It is considered that by incorporating the best aspects of each model, as was done for the knowledge lifecycle in Figure 3, an improved change maturity model could be developed. This will be the subject of a later paper. In the meantime the existing four models provides the framework for implementation.

3.2 The need for agility

Both Rogers model of technology diffusion and the capability maturity model are constructed as processes ending in full integration [50]. It is noted however from the literature that the time to achieve the integration presumed by Rogers' routinisation or level 5 maturity may be significant. Rogers himself suggests that the cultural change may take years [51] and a recent survey found that best practice can take an average of 27 months to move from one part of the organisation to another, even in the best of companies [52]. This indicates a lack of organisational agility to change in some instances. The turbulence of the business environment reduces the response-time available to firms to cope with change and thus creates the need for agility. The key drivers of this need for agility are to be found in a company's business environment, such as change in the marketplace, technology, and competition [53, 54].

Agility is suggested as more than just the ability to respond to change; it is the ability to incorporate change as a way of managing the business. This implies that agile organisations should make change be felt at all the organisational levels and activities; should possess a continuous and improved ability to predict changes and should be open learning organisations with an important external focus [55]. However the literature also notes that the more agile [56] an organisation is, the more dynamic and adaptive [57, 58], better able to operate profitably in a competitive environment of continuous and unpredictable change [59, 60], and more capable of gaining competitive advantage by intelligently, rapidly and proactively seizing opportunities and reacting to threats [61, 62]. It is contented that if the agility performance of an organisation could be better measured, management science could be applied to further improve that agility. This should enable more rapid maturity rate of knowledge change, creating greater capability for competitiveness.

Agile organisations are knowledge intensive organisations and are characterised because intellectual capital is the main factor of production and because they innovate in response to changing environments. These organisations create new knowledge bases by converting tacit and explicit and individual and collective knowledge reservoirs following a dynamic spiralling process [8]. Learning organisations should then be skilled at creating, acquiring and transferring knowledge and at modifying their behaviour to reflect new knowledge and insights [63].

4 PRACTICAL CASE STUDY IN MAINTAINING KNOWLEDGE ASSETS

A case study example is discussed here to demonstrate the practical concept of maintaining knowledge assets. A large coal mining company, operating a number of sites, is actively striving to better maintain its knowledge assets. The operational departments are nested deep within the corporate structure of the parent company as depicted in Figure 4. Obviously, in large organisations, the complexity of knowledge management is compounded by the amount of knowledge to be managed, the number of individuals involved, and the geographical separation / remoteness. Any change implementation issues with people, process or product also become more complex. By designing and refining effective and efficient solutions to issues that arise in such larger organisations, the solutions are then transferable to small implementations.

Figure 4: Organisational context of this case study

As part of its effort to continuously improve its business quality management systems, the Chief Executive Office supported the strategic vision for a company-wide Knowledge Management Policy (KMP). This KMP aims to optimise corporate information and knowledge management practices. The KMP was resourced and developed in response to recognition of the corporate value of information recorded within the organisation's knowledge, information and document assets. This case study discusses the practical aspects utilised to implement this directive at the 'operational' Department level, by considering the implementation issues with product, people and process, as follows:

- Product: To assist achieve the KMP strategic directive, the mine is currently implementing an Enterprise Content Management (ECM) System across all departments. The ECM system selected offers the technology to electronically store all documentation and retrieve items by searching for document meta-data or specific name strings, in a similar way to a Google search. Other products implemented to achieve this KM directive included a rationalisation of all archived documentation, with some obsolete documents being disposed of, and long term archives moved of site. The introduction of improved short-term knowledge storage systems in the product form of compactus and lockable cabinets. By creating a better space for document storage, the knowledge could be more appropriately sorted for application.
- Process: The site included process activities when planning the response to the KMP directive, acknowledging that the product (and product-focused training) would not be sufficient to achieve the corporate goals. A range of activities is being undertaken to assist the implementation process, including planning and offsetting issues before they became dysfunctional. This included regular communication to all employees involved through meeting presentations, e-mails, flyers, and one-one training. Another process initiative was the parallel implementation,

rather than sequential. This required project management techniques of scheduling, critical path analysis, and regular monitoring/review of progress.

- People: Observation over a period of 6 months showed that the implementation of this product is following an implementation maturity curve similar to that of Gaffe [47], with initial denial and resistance, to exploration and commitment. However there a few individuals – the early adopters (Rogers 1995) – who became mature users of the system with a minimum of fuss. Further process-focussed training and one-one consultations with stakeholders were applied to overcome or reduce resistance. Whilst the ECM tool enabled greater access to the organisational knowledge, it was observed that implementation was greatly influenced by the personal knowledge management disciplines / habits of individual people.

Overall, the practical experiences of this case have assisted in refining the emerging 'implementation maturity model' concept. The change curve was observed, and the methodologies developed to assisted offset and overcome the above people, process and product implementation issues before they became dysfunctional. This demonstrated the usefulness of the product-process-people framework. This case was presented here to demonstrate these practicalities in maintaining knowledge assets.

5 CONCLUSION

This paper discussed the concept of knowledge, its maintenance, the management systems that have evolved to assist, and the implementation of those systems. It is recognised from case study experience that it requires considerable skill to successfully implement a knowledge system change through to maturity. Implementation of a knowledge system change to maturity will assist organisations to maintain knowledge asset. Through management of the people, process and product issues, knowledge asset change was implemented, and overall knowledge system effectiveness increasing competitive capabilities of the business.

By pro-actively foreseeing and resolving any people, product and process implementation issues that arise, managers can increase the agility of any necessary change implementations. This not only makes the change more effective by being competitively mature earlier, it is also more efficient as resources are not wasted trying to push change against great resistance. The modern manager must appreciate the create/store/use/refine/dispose knowledge lifecycle, the maintenance requirements, the people, product and process issues that can arise, and the various stages of implementation maturity.

Ultimately, it is through better implementation of information and knowledge management systems that managers can better maintain their knowledge assets. The objective of this paper was to present knowledge management models that assist managers maintain organisational knowledge assets. These authors continue to research and refine their understanding of the implementation maturity model so as to provide change agents with useful tools to guide them through the implementation process. Through the development of the science of Knowledge Management, the challenge of change implementation becomes more manageable, and the client business becomes more competitive.

6 REFERENCES

[1] Toffler A. (1970) *Future Shock*. New York: Random House.

[2] Marshall A. (1965) *Principles of Economics*. London: Macmillan.

[3] Johanson U, Skoog, M. (2000) *Att mäta och styra verksamheten - modeller med fokus på icke-materiella resurser*. Uppsala: Uppsala Publishing House.

[4] IASC. (1989) International Accounting Standards Committee, Framework for the Preparation and Presentation of Financial Accounting Statements. *International Accounting Standards Committee*, 9.

[5] Edvinsson L, Malone, M.S. (1997) *Intellectual Capital: Realising Your Company's True Value by Finding its Hidden Brainpower*. New York, NY: Harper Business.

[6] Hussi T. (2004) Reconfiguring knowledge management - combining intellectual capital, intangible assets and knowledge creation. *Journal of Knowledge Management*, 8(2), 36-52.

[7] Ahonen G. (2000) Classification of Intangibles: Generative and commercially exploitable intangible assets. *Groupe HEC*, Jouy-en Josas, pp.206-13.

[8] Nonaka IaT, H. (1995) *The Knowledge Creating Company: How Japanese Companies Create the Dynamics of Innovation*. New York: Oxford University Press.

[9] Polanyi M. (1958) *Personal Knowledge; Towards a Post-Critical Philosophy*. London: Routledge.

[10] AS5037. (2003) *Knowledge Management Standard*. Standards Australia www.standards.com.au.

[11] Grayson CJ, O'Dell, C.S. (1998) Mining your hidden resources. *Across the Board*, 23-8.

[12] Davenport TaV, S. (2001) The rise of knowledge towards attention management. *Journal of Knowledge Management*, 5(3), 212-222.

[13] De Jarnett L. (1996) Knowledge - the latest thing Information Strategy. *The Executive's Journal*, 12, 3-5.

[14] APQC. (1996) American Productivity & Quality Centre, a nonprofit education and research organisation which fostered the creation of the Malcolm Baldridge National Quality Award.

[15] Miller W. (1999) Building the ultimate resource. *Management Review*, 42-5.

[16] Sveiby KE. (1997) *The New Organisational Wealth: Managing and Measuring Knowledge-Based Assets*. San Francisco: Berrett-Koehler Publisers Inc.

[17] AS15489. (2004) *Records Management Standard*. Standards Australia.

[18] Martin WJ. (2004) Demonstrating knowledge value: a broader perspective on metrics. *Journal of Intellectual Capital*, 5(1), 77-91.

[19] Quintas P, Lefrere, P., Jones, G. (1997) Knowledge management: a strategic agenda. *Journal of Long Range Planning*, 30(3), 385-91.

[20] Wiig KM. (1997) Knowledge management: where did it come from and where will it go? *Expert Systems with Applications*, 13(1), 1-14.

[21] Poynder R. (1998) Getting to the nuts and bolts of knowledge management. *Information World Review*, p20.

[22] Metaxiotis K. (2005) Exploring the world of knowledge management: agreements and disagreements in the academic/practitioner community. *Journal of Knowledge Management*, 9(2), 6-18.

[23] Senge PM. (1990) *The Fifth Discipline: The Art and Practice of the Learning Organisation*. New York: Doubleday.

[24] Quinn J. (1992) *Intelligent Enterprise: A Knowledge- and Service-Based Paradigm for Industry*. New York: The Free Press.

[25] Maier R, Remus, U. (2003) Implementing process-oriented knowledge management strategies. *Journal of Knowledge Management*, 7(4), 62-74.

[26] Pervan G, Ellison, R. (2003) Knowledge management practices and attitudes of IT departments: do they practise what they preach? *Journal of Information & Knowledge Management*, 2(1), 15-31.

[27] Liao S. (2003) Knowledge management technologies and applications - literature review from 1995 to 2002. *Expert Systems with Applications*, 25(2), 155-64.

[28] Wiig K. (2002) *New Generation Knowledge Management: What may we expect?* Arlington, TX: Knowledge Research Institute.

[29] Paraponaris C. (2003) Third generation R&D and strategies for knowledge management. *Journal of Knowledge Management*, 7(5), 96-106.

[30] Holmgren D. (2002) *Permaculture; Principles & pathways beyond sustainability*. Hepburn, Vic, Australia: Holmgren Design Services.

[31] Heeks A. (2000) *The Natural Advantage: an organic way to grow your business: 7 principles for high performance*. London: Nicholas Brealey Publishing.

[32] Hawking S. (1988) *A Brief History of Time*. Great Britain: Bantam Press.

[33] Riis JO. (1997) A situational maintenance model. *International Journal of Quality & Reliability Management*, 14(4), 349-66.

[34] Foucault M. (1980) *Power/Knowledge*. New York: Pantheon.

[35] Leonard D. (1999) *Wellsprings of Knowledge - Building and Sustaining the Sources of Innovation*. Boston, MA: Harvard Business School Press.

[36] Sohal A, Morrison, M. (1995) Is there a link between total quality management and learning organisations? *The TQM Magazine*, 7(3), 41-4.

[37] Sitkin SB, Sutcliffe, K.M, Schroeder, R.G. (1994) Distinguishing control from learning in total quality management: a contingency perspective. *Academy of Management Review*, 19(419-45).

[38] Water Vd. (2000) A maintenance model for quality management. *International Journal of Quality & Reliability Management*, 17(7), 756-770.

[39] Blanchard BS. (1986) *Logistic Engineering and Management*. NJ: Prentice-Hall.

[40] Moubray J. (1997) *Reliability Centred Maintenance*. Oxford: Butterworth Heinemann.

[41] Deming WE. (1986) *Out of the crisis*. Cambridge: MIT Center for Adv Eng Study.

[42] Sveiby K-E. (2001) Knowledge Management. *Journal of Intellectual Capital*, Volume 2(4), 344-358.

[43] Porter ME. (1980) *Competitive Strategy: techniques for analyzing industries and competitors*. New York: Free Press.

[44] Drucker PF. (1986) *Innovation and Entrepreneurship Practices and Principles*. New York: Harper Collins.

[45] Deming WE. (1982) *Quality, Productivity and Competitive Position*. Cambridge: Cambridge University Press.

[46] Rogers EM. (2003) *Diffusion of innovations*. New York: Free Press.

[47] Gaffe D. (1988) The Change Management Grid. *Training & Development Journal*, 42(4), 25-27.

[48] (SEI) SEI. (1995) Capability Maturity Model, In CM University (Ed) *Proceedings of*

[49] Satir V. (2003) A change management philosophy?

[50] Harris H & Ni J. (2004) Implementing the maintenance system: From human resistance to full integration, In (Ed) *Proceedings of International Conference of Maintenance Societies ICOMS,* Sydney ppPaper 28 p1-6

[51] Douthwaite B. (2002) *Enabling Technology*. London: Zed Books.

[52] O'Dell C & Grayson CJ. (1998) If Only We Knew What We Know: Identification and Transfer of Internal Best Practices. *California Management Review*, 40(3), 154-174.

[53] Zhang Z & Sharifi H. (2000) A methodology for achieving agility in manufacturing organisations. *International Journal of Operations & Production Management*, 20(4), 496-512.

[54] Sharifi H & Zhang Z. (2001) Agile manufacturing in practice - Application of a methodology. *International Journal of Operations & Production Management*, 21(5/6), 772-794.

[55] Srinidhi B. (1998) Needed: agile accounting to match agile organisations. *Agility & Global Competition*, 2(1), 41-55.

[56] Zhang Z & Sharifi H. (2001) Agile manufacturing in practice - Application of a methodology. *International Journal of Operations & Production Management*, 21(5/6), 772-794.

[57] Ashkenas R. (1995) *The boundaryless organization: breaking the chains of organizational structure*. San Francisco: Jossey-Bass Publishers.

[58] Ashkenas RN. (1995) *The boundaryless organization: breaking the chains of organizational structure*. San Francisco: Jossey-Bass Publishers.

[59] Goldman SNRPK. (1995) *Agile Competitors and Virtual Organizations: Strategies for Enriching the Customer*. New York: Van Nostrand Reinhold.

[60] Goldman SL, Nagel RN & Priess K. (1995) *Agile Competitors and Virtual Organizations: Strategies for Enriching the Customer*. New York: Van Nostrand Reinhold.

[61] J B. (2001) Developing manufacturing agility in SMEs. *International Journal of Technology Management*, (22), 28-54.

[62] Bessant J, et al. (2001) Developing manufacturing agility in SMEs. *International Journal of Technology Management*, 22, 28-54.

[63] Garvin DA. (1993) Building a learning organisation. *Harvard Business Review*, 71, 78-91.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support the University of South Australia provided to enable the research required to develop this paper. Also, the organisation in the case study (although not specifically named for confidentiality reasons) is acknowledged in its quest to maintain best practice knowledge management technologies to achieve business excellence. The insights in this paper are being further researched to assist CBD Consultancy continue to develop the science of Knowledge Management implementation, so as to build better business for clients.

DELIVERING OPERATIONAL SAFETY AND SERVICE AVAILABILITY THROUGH INNOVATIVE MAINTENANCE WORKSPLAN APPROACHES

John Ryan, [a] Mark Gilbert, [a] Max Bice [b] and Tim Morgan [c]

[a] *Systems Integrity, Safety Management Group, Airservices Australia, Locked Bag 747, Eagle Farm, QLd, 4009.*
[b] *Systems Integrity, Safety Management Group, Airservices Australia, GPO Box 367 Canberra ACT 2601*
[c] *Facilities Management Services, Technical & Asset Services Group, Airservices Australia, PO Box 1093 Tullamarine VIC 3043*

Abstract: Human intervention in system operation has the potential to introduce degradation in system performance, or, total failure of overall service delivery . The consequences of these fault or failure modes will vary according to the significance and configuration of the system, or service that has been interrupted by the planned intervention. Organisations must develop strategies and tools that provide engineering and maintenance staff, operators and decision makers with better information when planning to interrupt services for routine maintenance, project installation work, or configuration changes. Airservices Australia, as an internationally recognised Air Traffic Management Service Provider, has initiated an improved means of assessing and mitigating the risks associated with project and maintenance activities on Key Systems and Sites. These are defined as systems which, if interrupted, could cause a significant loss of service capability, seriously impairing operational safety. Our solution involved the development and support, in-house, of a successful Key Systems Works Planning process to capture the risks associated with the planned intervention, the mitigators required to reduce the identified risks to As Low As Reasonably Practicable (ALARP), the rationale for proceeding with the intervention and the associated communication requirements. The success of this program has been aided by the roll-out of a Human Error Management program.

Key Words: Systems Integrity, Safety Management Systems, Human Error Management, Works Plans, Maintenance, Operational Risk Assessment

1 BACKGROUND

For as long as industry has operated and maintained systems, plant and equipment, the very action of undertaking routine maintenance, configuration change, or, project installation, has resulted in the unplanned likelihood of degradation in system performance , or, at worst, total system failure. The consequences to production and safety are dependent on the operational significance and configuration of the system, or, service being interrupted. Industry can mitigate the outcomes of abnormal and degraded modes of operation by ensuring such activities are performed during complete plant shutdown and restart processes. This may not always be achievable, particularly in Air Traffic Management industry where it is common to be working on live systems, albeit with a backup, or, bypass system in operation.

A recent article [1] reported Britain's Air Accident Investigation Branch as stating "Maintenance errors were not the result of wilful negligence, or any desire to perform a less-than-satisfactory job, but the result of a combination of systemic issues that had increased the probability of an error being committed." British Airways (BA) took the safety report "very seriously" and had addressed the problems in its maintenance processes. "British Airways prides itself on safety and recognises that we are always ready to learn from incidents and encourage open transparent reporting," said the airline's head of safety, Captain Rod Young.

From a global perspective, we are witnessing an increase in system integration and operational inter-dependence, creating increased opportunities for common mode failures, yet, our strategies and tools for managing interruptions to systems and services can be seen to be lagging this technological change. In the challenge to strike the appropriate balance between production and safety, while still staying in business, organisations must continue to develop and implement strategies, tools and cultural change to enable engineering and maintenance staff, operators and asset owners to become informed decision makers with improved, appropriate and timely information when planning to interrupt services for planned maintenance, or, system configuration change. If we are to minimise the likelihood of operational failures when interrupting systems and services, we need to adopt a proactive approach,. thereby ensuring operational safety and system availability,

Early attempts at reducing operational safety incidents focused predominantly on the technological aspects, increasing reliability through improved engineering, driven from reactive analysis of accidents and incidents. Regulation of civil aviation defining 'what to do' and 'how to do it ' is essentially driven from the International Civil Aviation Organisation (ICAO)'s use of Standards and Recommended Practices (SARPs). In Australia the Civil Aviation Safety Authority (CASA), as the regulator, ensures international compliance by Airservices Australia to these SARPs . Substantial improvements in safety were achieved and continue to be achieved through this ongoing program

The implementation and application of a safety management system, utilising the safety case process, enabled Airservices Australia to systematically identify and manage hazards to reduce the level of risk to the National Airways System (NAS), through a 'systems' approach to safety. Certification and compliance to ISO9001, ISO14001 and CASR171/172/139 ensured sound systems' practices and procedures are adhered to, but these remained effectively a reactionary mechanism to safety improvement. What was needed was a proactive approach to safety culture through awareness and alertness to the effects of human error in the workplace [2,3].

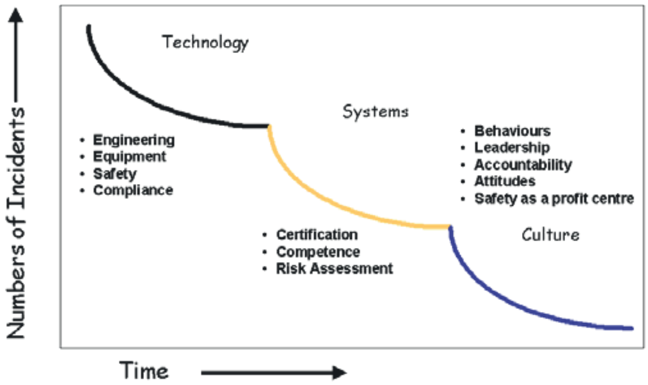


Figure 1 The Safety Management Evolution (11)

James Reason [4] describes a safety culture as the 'engine' that drives the organisation towards maximum attainable safety. Collectively it is the beliefs, attitudes and values of an organisation's approach towards safety. It includes the structures, practices, controls and policies the organisation possesses and employs, to achieve greater operational safety. Small incremental increases to safety have involved improvements to the safety culture of the organisation to better identify safety related hazards and risks, further reducers incidents, by instilling an 'intelligent wariness' [4], to expect the unexpected.

We were conscious of maintaining this "intelligent wariness" as a foundation for our safety improvement process. Airservices consequently developed a strong interest in human error identification, education and management in the engineering and maintenance group. Feedback from system performance reviews indicated that although we possessed highly skilled, knowledgeable and well trained staff, were a mature operator of a risk based Safety Management System, had embraced an integrated system safety philosophy, and had a comprehensive suite of support tools, we were incurring equipment failures that had the potential to impact on operational safety.

On several occasions Airservices Australia experienced a series of system failures arising from human errors by the engineering and maintenance group [5]. Subsequent investigation and analysis revealed a relative lack of human factors awareness, knowledge and capability within this group. This was identified as representing an ongoing risk to the safety and reliability of services. The Works Plan improvement program has been part of the safety culture journey, dovetailing with the roll-out of our engineering and maintenance Human Error Management program.

2 WORKS PLAN PROGRAM

2.1 Works Plan Concept

A Works Plan system facilitates the notification process and enables required actions when facilities/equipment forming part of the Australian national airways system are intended to undergo planned maintenance, configuration change, or, removed from operational service. The NAS includes facilities, equipment, [including Corporate Data Network facilities and privately owned facilities], used in the delivery of aeronautical telecommunications, radio navigation or air traffic services, as defined in Annex 10 and 11 to the ICAO Chicago convention.

"Works plans" are published descriptions of the intention to perform planned maintenance work. The nature of work and timing is proposed by technical staff and assessed and approved by operational staff. Once approved the execution of work is coordinated between Technical Customer Interface and Air Traffic Operations Managers.

The objectives of the original paper-based Works Plan system included:

- A mechanism for Operational, Technical and Engineering staff for planning the safe, efficient and effective removal and restoration of systems and facilities; this included works performed by internal and external service providers and where Airservices maintained assets for external owners, public and private sector
- Air Traffic Management (ATM) Facility Coordinators with advanced notice of equipment maintenance/configuration change activities that will affect their operations;
- The aviation industry to be advised of a temporary loss of a facility or a service, or, degraded level of service; (NOTAM action required);
- The ability for consultation among service users, prior to planned maintenance works;

- Formal advice of tasks that are considered high risk which may affect the provision of ATM services; and
- Increased level of visibility across the operational divisions of both scheduled and unscheduled maintenance, project and fault rectification activities.

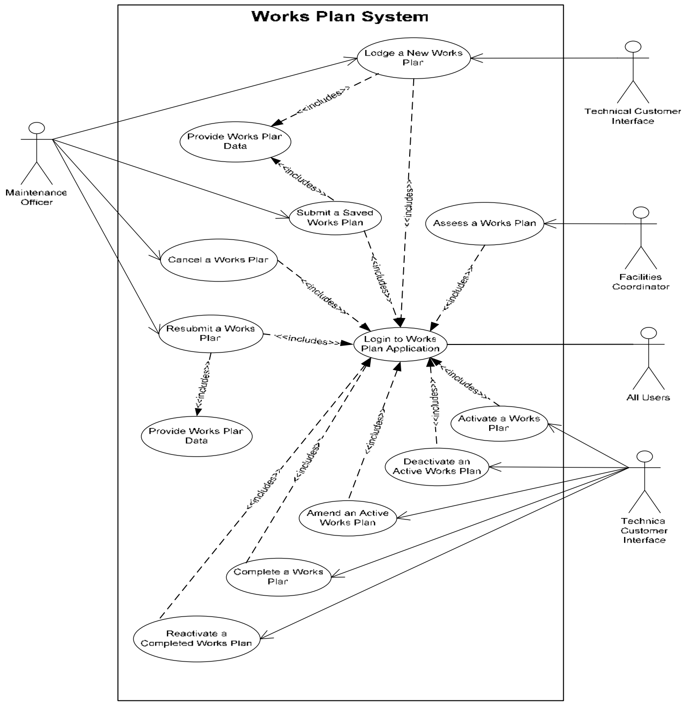


Figure 2 A Use Case Model that illustrates how various actors use the system by performing system functions (12)

The guiding principle for the use of works plans was defined as when routine maintenance, installation activity or testing will directly affect the ATC operating environment. Some examples included activity that:

- Required the operator to change their method of operation eg maintenance to a console where it required transfer of functionality to other consoles;
- Required the operator to be without a secondary or back up system or service for an extended period eg link path, secondary or tertiary frequency;
- May have a critical impact on multiple operational services; (testing of UPS static switches, satellite trunks)

Works Plans were not required for tasks on the airways system which were non intrusive, or, did not directly, or indirectly affect the facility user.

2.2 Intranet Works Plan

Performance audits revealed variations in use of the paper based system that had the potential to erode the safety requirements of the organisation in the face of constant socio-technological change, eg.

- Occasions where the works plan process was not used in accordance with agreed procedure
- Limitations of a paper based system in providing a uniform approach to daily management of Airways Systems works
- Inconsistency of information provided to decision makers eg- identifying hazards for risk determination
- Lack of visibility of both proposed & active works on an Flight Information Region, or, on a national basis
- Lack of Lessons learned Knowledge management engine

Airservices' solution involved implementing an intranet based Works Plan system to facilitate a uniform approach to the safe management of equipment maintenance activities. The system was developed "in-house" as a replacement for a paper based equipment maintenance process, previously used in isolation across all locations in Australia. This development strategy was driven by the lack of a commercially available "off the shelf" product at the time of need, significant internal programming knowledge and skills and a desire to maintain the safety and security around the propriety operational knowledge of the NAS.

Since the initial release of the "live" system, there had been many enhancements to the Works Plan application, in order to improve functionality and meet vital emerging safety and operational requirements. These were identified through regular feedback as well as planned consultations with stakeholders, including, asset owners, operators, maintainers, engineering and operational authorities.

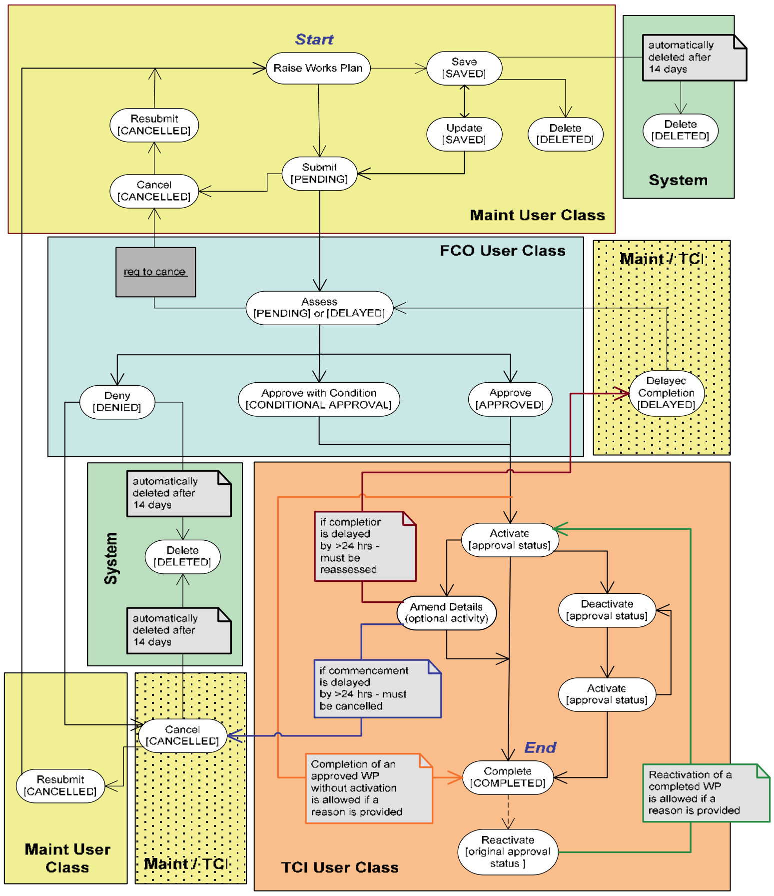


Figure 3 Illustration of the actions various user classes perform during the lifecycle of a works plan. (12)

Although the modifications have served the purpose at the time of implementation, the application had become inadequate in the changing operational environment. From a socio-technological perspective the existing system was deemed inefficient and expensive to maintain.

2.3 Enhanced Works Plan

The functionality of the existing intranet Works Plan application required improvement in order to maintain organisational missions of 'Keeping Safety First' and adopting world's best operational practices and safety requirements. A project was established to implement an Enhanced Intranet based Works Planning tool that would provide strategic and tactical planning capability for actual and potential service interruptions. A key requirement that emerged was the visibility of proposed maintenance, project activities, and system operational status with pre-existing conditions, such as degraded modes and known fault conditions. This required a mapping of systems to services, in a way that clearly showed the functional interrelationships. With this information, it would be possible to apply a risk based approach to service interruption based upon AS4360 – 2004, utilising the knowledge and expertise of the respective domain experts to identify and assess the technical and operational risks associated with maintenance activities and their potential impact on service levels. The Enhanced intranet based Works Planning Tool would provide a mechanism for effective communication and liaison between operational, engineering, technical work areas and relevant stakeholders. The solution would facilitate a means to capture knowledge of known adverse system interactions and interdependencies to minimise unintended adverse outcomes.

Initially there were difficulties enacting the mapping of systems to services, given the complexity of the operational environment. There was and still is, a reliance on the tacit knowledge of engineering and operational authorities to draw together the "big picture' of the operating environment when works plans are lodged for approval to proceed. With the continued use of the application, this tacit knowledge is being converted to explicit knowledge, retained within the application and through interfaces to SAP and the Document Management System.

In summary, the immediate benefits of the Enhanced Intranet Works Plan redevelopment project have included:

- Improved visibility of proposed maintenance activities to alert operational and maintenance staff on planned concurrent site activity and activity at other dependent sites enabling safe and efficient removal and restoration of systems;

- The identification of Key Systems providing a greater understanding of the technical and operational risks associated with system and site maintenance; this has lead to the development of Key Systems Works Plans and coordination with the Operational Risk Assessments (ORA). Figure 4 details the structure for ORA development. The top level encapsulates the delivery of an integrated service to the aviation industry. Underpinning the delivered services are airways systems and service delivery functions. Service delivery functions include system operation, maintenance and life cycle management

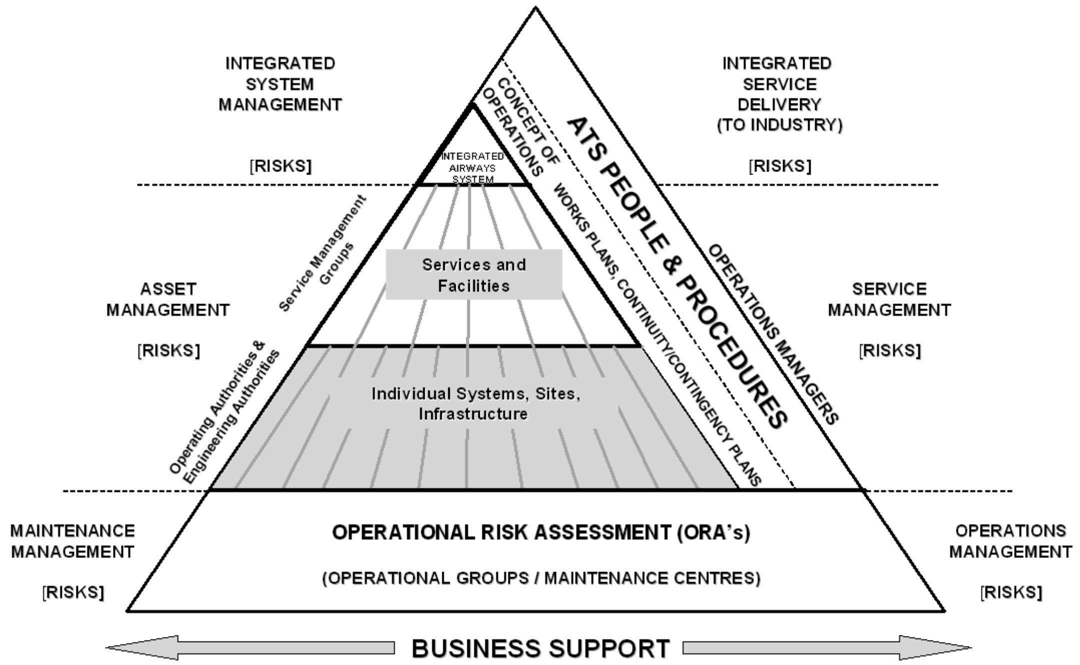


Figure 4 Operational Risk Assessment Pyramid (12)

- Data synchronisation with SAP PM to provide maintenance staff access to up to date information on equipment, systems and services when planning maintenance activities.
- Regular data uploads from SAP PM on faulted services to alert maintenance staff on known fault conditions when planning equipment maintenance activities;
- Links from Works Plan system to Key Systems Works Plan documentation and online system overview site diagrams to enable maintenance and operational staff make informed decisions relating to activities that may interrupt service delivery; and
- Enhanced screen layouts and modifications to existing system functionality greatly improved the efficiency and usability of the application; User training courses tailored for different user groups
- An Airservices wide, user friendly integrated approach, with a level of accountability for works plan activities across Operational, Engineering and Technical domains

2.4 Key Systems Works Plans (KSWP)

As part of our continual safety improvement program, Airservices Australia has developed an improved means of assessing the risks associated with periodic maintenance activities on Key Systems, determining appropriate mitigators to reduce the risk to ALARP, the rationale for proceeding with the maintenance and the associated internal and external communication arrangements. Key Systems are defined as those systems which, if interrupted, could cause a significant loss of service capability. That is, services that are critical to the organisation's mission. Soon after commencing this activity the group realised that periodic maintenance at Key Sites could also cause a significant loss of service capability. Key Sites are those sites which, if disrupted, could cause a significant loss of service capability. Key sites include those sites upon which multiple other sites have a high dependency for the integrity of service delivery, or those which support mission critical systems. These may include sites that contain a significant number of facilities, or that support a key system eg. A hub link site that feeds a number of other sites which, individually, present a low risk to service delivery, but, if all sites were not available on-line due to the failure of the hub link site, the risk to service delivery is greatly increased. The hub link site might then be considered a key site. If one of these sites supported a key system it too could be considered a key site.

The KSWP program was facilitated by the development of Key System Risk Management Reviews. These reviews recorded the outputs of a structured process that captured the risks associated engineering and maintenance activities at Key sites. Awareness by the facility co-ordinators of the growth in system integration and operational inter-dependence, with increased opportunities for common mode failures was an early milestone. An immediate need was flagged by the engineering and operational authorities to prepare simplified block diagrams, from the abundance of existing detailed schematics. These would be available to the Technical Customer Interface and Air Traffic Operations Managers, as an aid to clarification of the operational impact of intrusive maintenance on Key Sites and services before a decision was made to proceed or defer maintenance activity. An example simplifying the power supply to a major control centre is shown in Figure 5 .

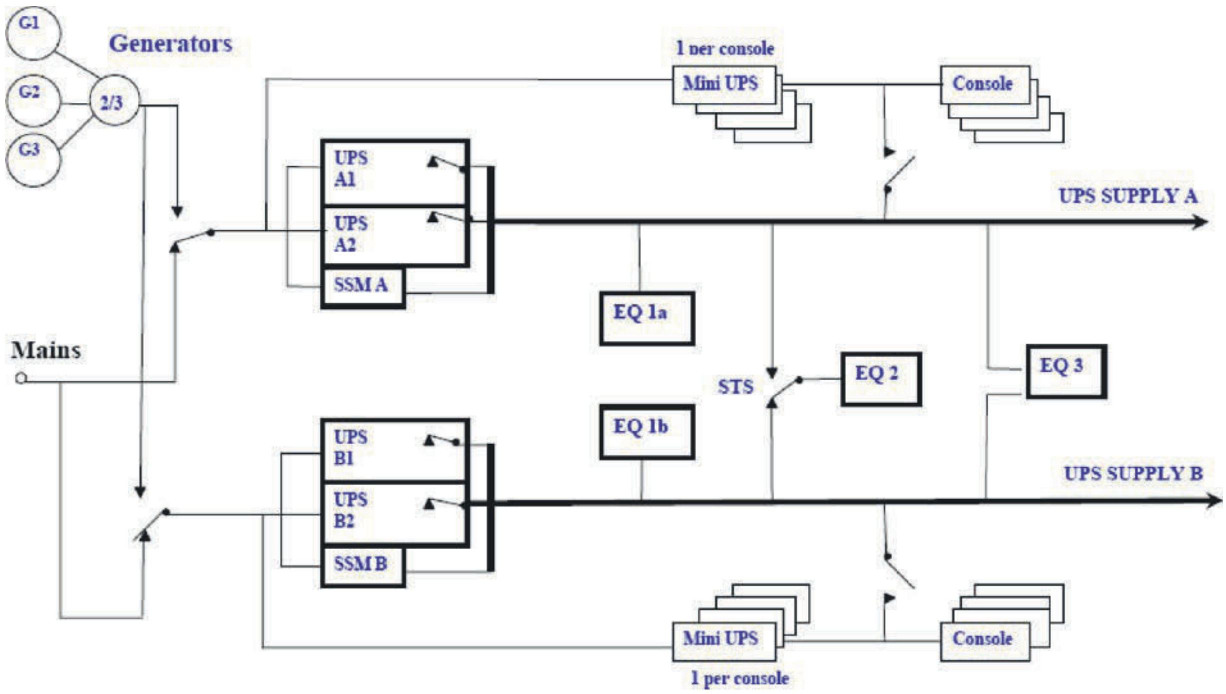


Figure 5 Basic UPS configuration for Major Air Traffic Control Centre (12)

Numerous benefits have flowed from this work, some of which include:

- Greater understanding of the technical and operational risks associated with Key System/Site maintenance
- Enhanced defences in reducing both the likelihood of failure and operational preparedness should a failure occur
- Improved communications with operations leading to improved decision making and a streamlined approval process for key system maintenance
- Improved definition of the system baseline uncovering potential improvements in system configuration to cater for system degradation

2.5 Works Plan categorisation

Key System Works Plans are included in Works Plan method, and are the top level category of Works Plan. Other categories of works plans would be required, with categories such as:

- Moderate (Written Consultation)
- Minor (Verbal Consultation)
- Nil Ops Consultation Required (eg data backups, non-operational equipment)

Decision support criteria for operations and maintenance decision makers have been developed for each category, based on the "significance" of the proposed works. These criteria can be expressed in descriptive terms, using examples of types of systems to be maintained, or, using the consequence categories, or a combination of both.

2.6 Threat Alert Management

The Threat Alert Management concept is meant to forcefully interrupt the "chain of causation" leading to an operational safety incident/accident by alerting engineering, maintenance and operators to lessons learned and potential hazards in the execution of the planned activities[6]. In figure 6 we see an adaptation of James Reason's accident causation/ "Swiss Cheese"

Model [4] that serves to highlight the process of error creation, capture and management, or, if all the “holes line up”, system safety failure.

Table 1 Criteria for Works Plans categories

Category	Consultation	Description	Consequence Criteria
Key System	Written	Potential for major loss of service provision, such as multiple frequency failure, multiple radar/display failure or a combination of system failures leading to a major degradation of service	The ability to maintain an air traffic service may be severely compromised within one or more airspace sectors without warning for a significant time.
Moderate	Written	Potential for moderate degradation of service provision (but no loss of service), handled through normal contingency arrangements	The ability to maintain an air traffic service may be impaired within one or more airspace sectors without warning for a significant time.
Minor	Verbal	Monitored by Ops SS	No affect on the ability to maintain an air traffic service in the short term, but the situation needs to be monitored and reviewed for the need to apply some form of contingency separation measures if the condition prevails.
Insignificant	Nil Ops Consultation	Technical Notification only (TCI monitoring), non-NAS related maintenance	No possible affect on aircraft or ATC operations

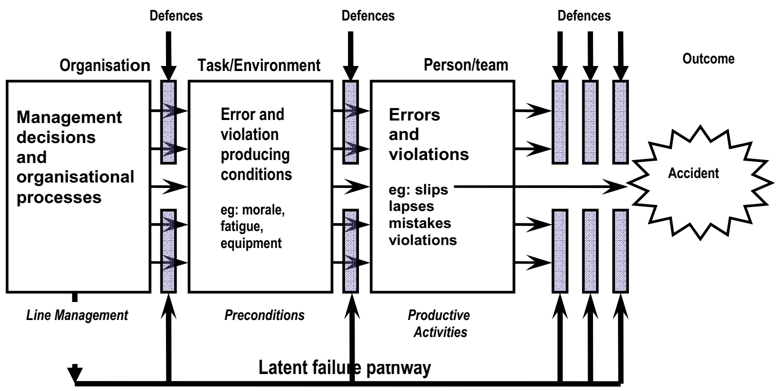


Figure 6 Adapted from Reason Model of Accident Causation [4]

The Threat Alert Management concept involves identifying and capturing potential threats to system integrity, that, if not properly considered, could adversely influence the outcome of a particular maintenance or project activity. Threats are the “ Traffic Lights” of potential hazards. Threats can, if not properly identified and considered, add to workload, distract, or force staff to do something different to the way one would normally act. These situations can increase the likelihood of human errors and could lead to undesirable outcomes. Threats to Works Planning include:

- Inter and intra-site dependencies (work at other “dependent” sites)
- Concurrent works (other active works plans on related facilities)
- Other works that, if combined with planned works could have an undesirable outcome (including external or contractor works)
- Air Traffic congestion; weather conditions
- Technical and operational staffing constraints
- Known equipment / system degradation or performance issue; lack of operational system bypass
- Undesirable times of day for particular work
- Duration of work

- Free lessons and near misses that capture relevant "things to watch-out for"

A threat may prevent engineers, maintainers, or, operators from planning and executing their job correctly. It may have the potential to impact on safety or efficiency. Maintenance and engineering staff, together in consultation with operations staff have the knowledge, skills and abilities to identify the types of threats. The Threat Alert Management Concept provides an approach that would improve situational awareness for both operations and maintenance staff, enable improved decision making for system or service interruptions, and reduce the likelihood of undesirable outcomes. Threat management will be incorporated into the Works Planning process, creating a database for capturing threats enabling the information to become visible for those involved in the Works Planning process.

2.7 Just Culture and the Blindness to Unidentified Risks Norms and Threats (BURNT)

Our ability to minimize, capture and mitigate operational error relies on our personal competencies, supported by an external systems and process that are error tolerant. A baseline essential requirement is an effective system of reporting errors that is supported through an organizational commitment to "Just Culture". Airservices commenced the rollout of this program in 2004, with firm commitment from our CEO and Executive and commenced development of a Human Error Management Program (HFHRO), in conjunction with an international expert in the field [6]. important goals of this program were to:

- enhance Airservices Australia's safety culture by establishing a "Just Culture", where staff willingly and openly report human errors, near misses and free lessons, without fear, or, intimidation and
- reduce instances of human error, by promoting an understanding of contributors that lead to making errors, and an awareness of the safety nets that can be applied to negate these contributors.

Feedback from staff who participated in the program has proven to be extremely valuable in identifying issues which may otherwise have remained undetected in the workplace. In recognising that we were on a journey, Airservices Australia's management subsequently endorsed the BURNT (Blindness to Unidentified Risks, Norms and Threats) training program.

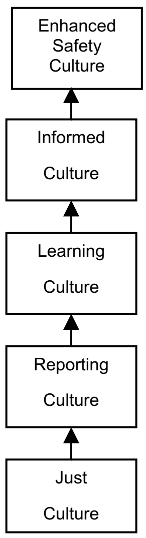


Figure 7 Managing Human Error through enhanced Safety Culture (12)

BURNT provided participants with a range of skills to identify norms and risks in the work place. There are many healthy norms in the workplace that contribute positively towards safety but are undocumented and not well known. The program provided staff with the techniques to identify these norms, so the benefits could be more widely distributed. Conversely, there may be norms that have a negative impact on safety and system integrity and therefore present a risk to staff, customers and our business. Through this program, these norms continued to be identified for appropriate attention by staff and management, working together. The overall theme being one of risk management and awareness.

BURNT workshops were of two days duration and presented in our workplace. The workshops were highly interactive and examined a number of case studies from multi-national High Reliability Organisations. An Operational Risk Assessment (ORA) concluded the workshops. Here, issues in our workplace that represent threats or opportunities to the way we do business are identified, discussed and actions planned, to ensure that the maximum benefits are obtained from the program The enhanced Works Planning process was reinforced and tested.

Our organisation has utilised information gathered from the BURNT Workshops to increase our understanding of the risks to our business operations. This process will continue to enhance our existing internal risk management processes and the growth of the safety Knowledge management system that is part of the Enhanced Works Plan tool.

In addition, the recent application of a tailored Human Error Investigation and reporting system, based on the error producing drivers, identified as local factors and maintenance factors as shown in figure 8 [5], has emerged as a significant diagnostic intervention in the journey to an enhanced safety culture

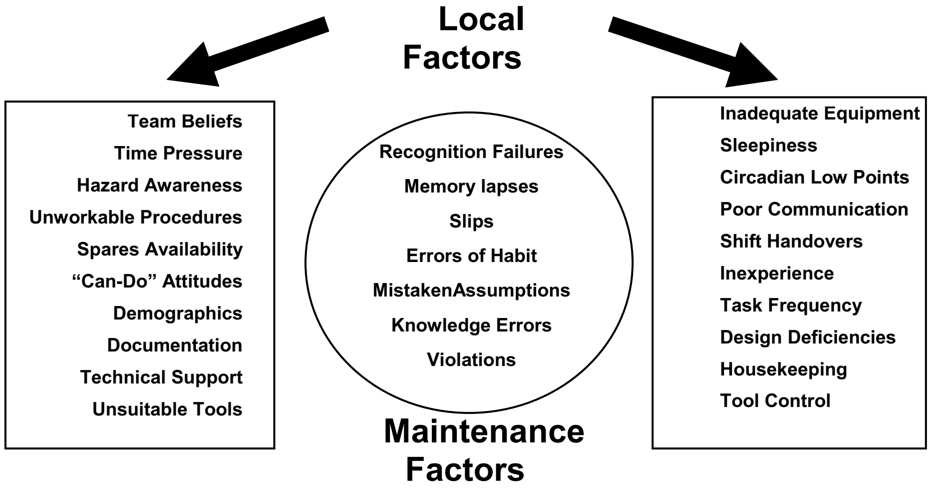


Figure 8 Adapted from Reason, J & Hobbs, [5] pg 63

3 NEXT STEPS IN THE SAFETY JOURNEY

In the daily challenge to balance production and safety imperatives, organisations are focussed on navigating the rocky path between catastrophy and bankruptcy [4,7]. In order to survive and grow, we continue to seek out strategies for best practice to enable optimal performance while ensuring safety of our employees and the travelling public.

Maintenance anomalies constitute one of the most common failures leading to system breakdowns, outages and accidents in a range of industries. Errors can be seen as mismatches between the demands of the task and the capabilities and limitations of the human. With appropriately targeted interventions, the probability of error can be reduced [8]. However, even in an ideal environment, a low background rate of human error is inevitable. For this reason, a maintenance error management program not only requires interventions to reduce the probability of error, but also requires the organization to continually increase the resilience of the system in the face of error [5,9,10] Airservices Australia has identified the following improvements in its safety management journey:

Enhanced Key Systems and Sites Works Plans

- Continued review and Integration of KSWP's into business operations
- Review service continuity consequence criteria against both recovery time and the interruption duration
- Continued Development of the threat and error concept; technical benefits/concurrent site activity alerting
- Existing system faults degraded/abnormal operation modes visibility in works plans
- Verify and validate historical reasons for customer rejection of works plans
- Verify and validate coordination agreements with external agencies and service providers

Implement operational safety measures into outage tasks

- Identifying tasks with high incidence of error
- Dissemination of best practices and safety guidelines to staff
- Analysing task management to determine best practices
- Continue development and application of Human Error Management Programs

Identifying and defining key areas of focus for employee performance and safety

- Establishing an ongoing (re)training programme to maximise operational efficiency and minimise incidents; Decreased incidence of injury and loss of workforce; increased equipment reliability;
- Asset manager, operator and maintainer training needs analysis

4 CONCLUSION

An intranet Works Plan system facilitates the notification process and enables required actions when facilities/equipment forming part of the Australian national airways system are intended to undergo planned maintenance, configuration change, or, removed from operational service. Works plans are published descriptions of the intention to perform planned maintenance work. The nature of work and timing is proposed by technical staff and assessed and approved by operational staff. Once approved, the execution of work is coordinated between Technical Customer Interface and Operations Manager*s*

Improvements in the delivery of operational safety and service availability can be achieved through an enhanced worksplan process. The intranet worksplan process facilitates the notification of intended work on the Airservices National Airways System including planned maintenance, configuration changes, projects or any work which may affect an operational service. Worksplans are categorised into Key System\Site, Moderate, and Minor. Improved decision making can be achieved by providing improved understanding of service to system mapping, analysing and prioritising the risks of human interaction, and determining contingency procedures on key systems and at key sites. The benefits are reduced unplanned service interruptions and in the event of a failure, a reduced impact on key systems and sites.

The worksplan process on systems with the potential for a moderate degradation of service provision include an analysis of the affect on operations and a risk assessment of the potential of an unintended outcome from the work.

5 REFERENCES

[1] Creedy, S. (2006) Cracks appear in British airline safety, , The Australian, January 02, pg 6

[2] Guldenmund, F.W. (2000) The nature of safety culture: a review of theory and research. In Safety Science, 34, 215-257. Elsevier

[3] Hudson, P. (1999) Safety culture –the way ahead? Theory and practical principles. In Proceedings – International Aviation Safety Management Conference. pp. 93 – 102: Perth, WA. Edith Cowan University.

[4] Reason, J (1997) Managing the risks of organisational accidents. UK. Ashgate Publishing

[5] Reason J. and Hobbs A. (2003) Managing Maintenance Error – A Practical Guide. Aldershot, England: Ashgate Publishing Company

[6] Marx, D (2001) Patient Safety and the "Just Culture": A Primer for Health Care Executives Medical Event Reporting System – Transfusion Medicine (MERS-TM), , David Marx Consulting, "Prepared by David Marx, JD, for Columbia University under a grant provided by the National Heart, Lung, and Blood Institute "

[7] Westrum, R. & Adamski, A.J. (1999) Organisational factors associated with safety mission success in aviation. In Handbook of Aviation Human Factors, Garland, D.J. & Wise, J.A. & Hopkin, V.D.: USA. Lawrence Erlbaum Associates.

[8] Airservices Australia (2005) Human Error Investigation. Canberra, ACT.

[9] Cacciabue, P.C. (2004) Human error risk management for engineering systems: a methodology for design, safety assessment, accident investigation and training: In Reliability Engineering & System Safety, 83 , 229-240: Elsevier.

[10] Helmreich, R.L. & Merritt, C. (2001) Culture at work in aviation and medicine. UK. Ashgate Publishing.

[11] Redmill, F Anderson, T (1993) Safety-critical Systems: Current issues, techniques and standards. Chapman & Hall, London. ISBN: 0-412-54820-8

[12] Airservices Australia (2005) Key Systems Worksplans, Internal Reports

Acknowledgments

Airservices Australia: Mr Chris Howell, Manager Services Integrity for supporting this paper; Safety Management Group team members for their continued efforts to improve our safety change management processes; Information Management Services Group for developing IT product solutions.

RFID FOR ASSET MANAGEMENT: A FACT OR FAD

Abrar Haider [a] and Andy Koronios [b]

[a] *School of Computer and Information Science, University of South Australia, Mawson Lakes, Adelaide 5095, Australia*

.[b] *School of Computer and Information Science, University of South Australia, Mawson Lakes, Adelaide 5095, Australia.*

Abstract: Radio frequency identification (RFID) is a powerful data capturing technology that provides electronic identification, tracking, and storing of information contained on tags attached to or embedded in items, such as products, cases, or pallets. However, main trust of businesses in adopting this technology has been on logistics and warehouse management. RFID has particular significance for industries like defence where inventory consists of a variety of items; in order to alleviate capacity constraints, flow of materials in reverse direction is high; supply and demands factors are not as stable as compared to some other supply chains; and geographic end points of supply chain are not stable. RFID technology takes workforce mobility to a higher level by enabling eliminating the human computer interaction in tasks such as inventory tracking and management, and even condition monitoring, thereby allowing for reduction in data entry errors, speed of data flow, and real time access to information. This paper presents the results of an evaluation study undertaken for the Cooperative Research Centre of Engineering Asset Management to explore the efficiency of RFID technology for Asset Management. It identifies the major initiatives taken by different industries to use the technology; discusses current trends; and provides an evaluation of three technologies used to investigate the potential and limitation of RFID technologies. Drawing upon the conclusion it provides further insights into areas where RFID technologies could be used in the area of asset management.

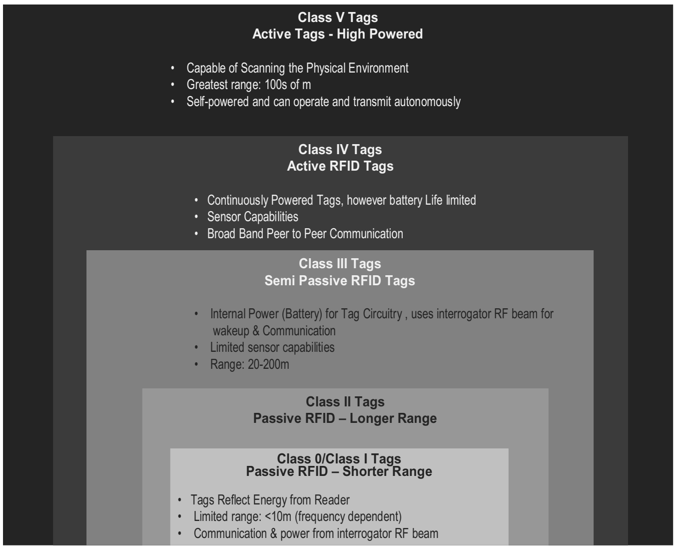
Key Words: Asset Management, Asset Maintenance, Performance Evaluation, Information Systems.

1 INTRODUCTION

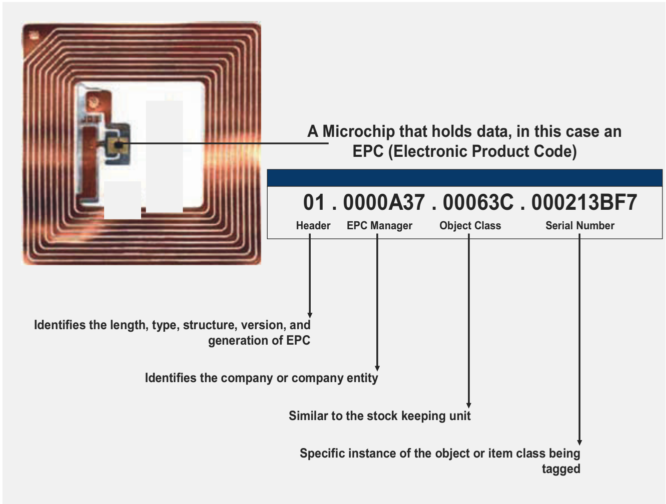
Radio Frequency Identification (RFID) has gained enormous popularity in the recent past, even though the technology has been around for considerable period of time. Reasons for this trend is the ease with which RFID technology provides for item tracking and thereby inventory management, logistics management, and supply chain management. Wal-Mart’s and US department of defence have been leading the race in this regard, and in so doing are realising a novel relationship between items and information. Major advantage of RFID technology over bar code technology is that with RFID technology information exclusive to the item travels with the item throughout the life cycle of the item. For example, at any time it is possible to find out the current location of the item; its origin, such as supplier and date of manufacture; its existing use and intended use; and its future destination. This functionality provides for real time visibility of end to end supply chain. RFID has particular significance for defence supply chains where inventory consist of a variety of items; flow of inventory from consumer to supplier is high; and supply and demands factors in this situation are not as stable as compared to some other supply chains, and geographic end points of this supply chain are continuously moving. However, this is just one area where RFID technology offers potential benefits; in fact it has the potential to address diversified issues concerning asset health management. Nonetheless, the proliferation of RFID technology is subject to the development of technology to overcome some of the issues such as, lack of standards, cost, data/infrastructure/systems implementation issues, data sharing and security. This paper presents an assessment of RFID technology, its applications and implications, and future research directions with particular significance to engineering asset management. Although RFID technology is convincing, as yet it is far from being pervasive. Nevertheless, we conclude that this technology is here to stay and its capabilities provide excellent grounds for various wireless applications in asset management.

2 UNCOVERING RFID TECHNOLOGY

RFID technology consists of three components, an RFID tag, a reader, and a Tag/Reader management systems or a middleware [1]. RFID tags are made up of a small microcontroller and antenna available in many different packages. They provide a contact free form of identification through the use of radio frequencies. Each tag has an electronic product code (EPC) or an identification number embedded into the tag microcontroller that is used to uniquely identify each tag, which can also be termed as the RFID’s version of a bar code [2]. When an RFID tag is placed close enough to the reader it is powered up through a magnetic field emitted by the reader thus powering the microcontroller of the tag, such that it transmits the EPC to the reader. RFID tags do not require line of sight between tags and readers for them to be detected and therefore make it possible for tags attached to items to be identified from a single point [3] [4]. An RFID reader is a mobile or fixed device, comes in a variety of shapes such as hand held or fixed mount, and emits electro magnetic waves. An RFID reader activates the tags by sending tags out encoded interrogation signals. This signal creates a magnetic filed that the tag uses to power itself, the tag then sends its EPC to the reader. The EPC is decoded by the reader and sent to the middleware. An RFID middleware represents the hardware and software that is necessary to process data coming from the tags, coordinate communication to and from multiple readers, and sends the captured information to legacy systems. It also segregates the vast amount of data before sending it onwards to legacy systems. In simpler terms it works as a traffic regulator and to some extent it also works a communication moderator. Nevertheless, unlike bar codes that are only capable of identifying a class of items, EPC identifies each item uniquely. The EPC identifies the manufacturer, product, version, and serial number or a product or item. There are two types of EPCs in common use, one that supports 64 bit on board EPCs and second the ones that support 96 bit EPCs. A 96-bit EPC provides unique identifiers for 268 million objects [2]. Figure 1b illustrates the taxonomy of an EPC.



a. RFID Tag Classification



b. Electronic Product Code

Figure 1

RFID tags can be active, passive, or semi-passive tags, and are classified into five classes according to their capabilities and functionalities (Figure 1a). However, there are commonly referred to as either passive or active. Passive tags are interrogated by the reader for information contained in them, whereas active tags themselves transmit the data streams to the reader, hence the name passive and active. Passive tags are powered via the reader thus they do not require batteries to operate, whereas active tags have an onboard power. Passive tags can be detected from a distance ranging from a few centimetres to a few metres, whereas active tags, on the other hand, have an on-board battery, and therefore, have a far longer read/ write range and memory size. Class 1 and 2 tags are the most common and can be easily found in operation today. These tags are mostly used for item tracking and inventory management purposes. Class 3 tags are semi passive, which means that they are on board power available, but that power needs to be activated by the magnetic filed that the sensing beam of the RFID reader creates. They have relatively smaller read range, which means that the reader has to be placed in close proximity to the tagged item. Class 4 and 5 tags have on board continuous power available (usually a life span of 10 years) as they are fitted with a battery, in most cases have a microcontroller available on board, support on board memory, and also have sensing capabilities. These tags are usually able to sense the temperature or light. Table 1 summarises typical characteristics of active and passive tags.

Table 1

Typical Characteristics of RFID Tags

	Passive tags	**Semi passive tags**	**Active tags**
Power supply	external (from reader)	internal battery	internal battery
Read range	up to 20 feet	up to 100 feet	Up to 750 feet
Type of memory	mostly read-only	read-write	read-write
Cost	$.20 (US) to several dollars	$2 (US) to $10 (US)	$20 (US) or more
Life of tag	up to 20 years	2 to 7 years	5 to 10 years

Source: [5]

Active tags have external interface available that allows for interfacing additional sensors and devices. On the other hand, major advantage of the passive tags is their small size; consequently they can be used for many applications where size and weight are limiting factors. Extra features on a tag, however, do increase the size and cost of each individual unit. Table 2 summarises the characteristics of passive tags and indicates application areas for each segment of tags.

Table 2

Common RFID Operating Frequencies for Passive Tags

	Frequency	Typical read range and rate	Examples of use
Low frequency	125 KHz	~1.5 feet; low reading speed	Access control, tracking, point of sale applications
High frequency	13.56 MHz	~3 feet; medium reading speed	Access control, smart cards, item-level tracking
Ultrahigh frequency	860-930	MHz up to 15 feet; high reading speed	Pallet tracking, supply chain Management
Microwave frequency	2.45/5.8 GHz	~3 feet; high reading speed	Supply chain management

Source: [6]

3 RFID APPLICATIONS

There are numerous applications of RFID technology currently being developed and a plethora of applications are envisaged by prospective users. At the moment major advantages that businesses gain from RFID are item tracking, warehouse management, reduced obsolete inventory, reduced material handling costs, supplier identification, authenticity of items, end to end visibility of materials flow, and supply and demand management support. Among these are applications like tracking a box, web luggage, smart shelves for pharmaceutical products, smart toolbox, and smart box etc [7], [8], [9], [10], [11]. With regards to asset management, there are a range of applications for RFID technology spanning asset management particularly maintenance management. Active RFID technology can be of significant use for condition monitoring of an asset; the sensing capability of RFID technology is further discussed later in the paper. The tracking capability of RFID could be used for tracking assets such as railway tracks, and pipelines. Table 5 presents the some of the RFID related initiatives under way in different agencies.

Table 3

Reported or Planned Use of RFID Technology

Industry/Service	Application
Defence	Logistics support; Tracking shipments ; Tracking and identification; Monitoring weapons
Energy	Detection of prohibited articles; Tracking the movement of materials
Health and Human Services	Physical access control
Homeland Security /Immigration and customs	Border control location system; Smart containers; Tracking and identification of assets Tracking and identification of baggage on flights
Law and order	Tracking offenders and prisoners
Transportation	Fleet management, Electronic screening
Finance and Treasury	Physical and logical access control; Records management (tracking documents)
Environment Protection	Tracking hazardous material
General Service Administration	Tracking and routing carriers along conveyor lines; Identification of contents of shipments; Warehouse management

Source [12]

4 MAKING USE OF THE UNIQUENESS

Figure 2 illustrates a framework that was used for developing different applications to investigate the potential of RFID technology for asset management. In this framework each item is tagged with an RFID tag; either active or passive, depending upon the nature of application. When the tag comes near the reader, the radio frequency field generated by the reader powers up the tag and causes it to continuously transmit its EPC by pulsing the radio frequency. The reader passes the EPC to the middleware, or what is also termed as a savant. A savant acts as a buffer between the reader and other organisational information systems, and consists of various modules or sub programs with each module performing specific functions.

The savant remains connected to the readers, and behave as a router of the RFID network with the primary functions of EPC related data smoothing, data forwarding and data storage; along with reader coordination, and task and event management. These savants use algorithms that take care of the reader collision, such that each tag is identified uniquely and is read each time it is attempted to be read.

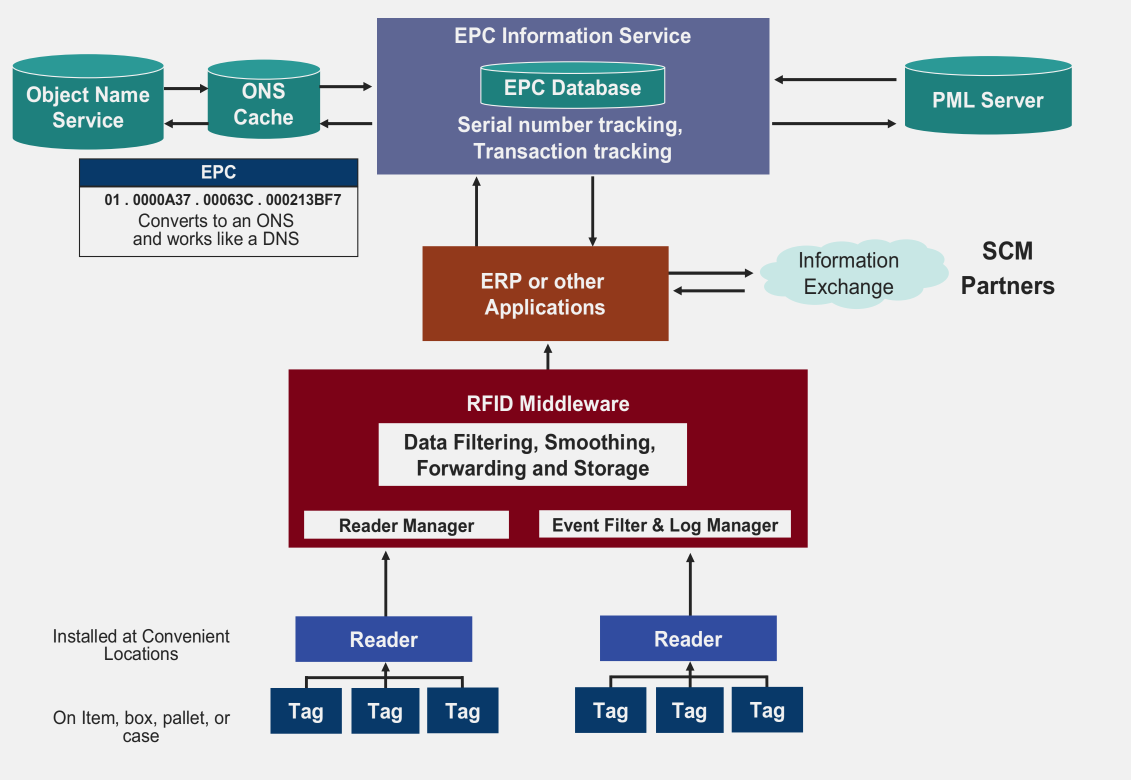


Figure 2 –Framework for RFID Based Prototype Applications for Asset Management

Since the EPC is the only information stored on the tag, it has to be used in such a way that it provides additional information about the item that the EPC is attached to. The savant interprets the EPC into a unique address of an object naming service (ONS), which points to a physical mark-up language (PML) server. PML is a version of the popular XML meta-data language. The server stores PML files that contain information about the item that the RFID tag is attached to. The whoel process works like a user access a website on the Internet, where a user types in the unique url (in this case the EPC), this url is interpreted into a DNS, which points to the service that is hosting the website pages and associated information. Using this framework, any amount of information exclusive to an item can be stored and retrieved in real time.

5 EVALUATION OF RFID TECHNOLOGIES FOR ASSET MANAGEMENT

In our evaluation we examined three RFID Technologies, for configuration management, item tracking, and for condition monitoring. We chose Texas Instruments Midrange Evaluation kit and Chipcon CC1010, since both these technologies offer valuation packages which include all the hardware and software required to begin developing specific applications. This also allowed us to explore both the passive and active tags. Even though the scope of this evolution was provide the proof of concept of RFID utilisation for configuration management, nevertheless, we were able to demonstrate tracking capabilities of RFID technology as well as sensing capabilities. Examination of active tags led us to believe that RFID technology could be used of a much enhanced scope of applications in engineering asset management.

5.1 RFID for Configuration Management

The Texas Instrument midrange evaluation kit is equipped with one reader and a selection of Tags in different packages. It operates at 13.56MHz capable of reading tags at a range of 7 inches from the reader. The reader unit is a similar size to a paperback novel (13cm*3.5cm*18cm), the unit is encased in a black plastic box which roughly doubles the necessary size of the unit. A variety of tags are supplied with the reader unit they can be categorized into two different types which come in different sizes. The Tag-it transponder inlays with 256 bits user programmable data, the estimated data retention time is 10 years and can be programmed typically 100 000 times. The evaluation kit also comes with Tag-it ISO transponder inlays also available in a range of different sizes the main difference is they have 2000 bits of programmable memory and a larger unique identification number for future expansion. The inlay style tags are a cost effective solution for tracking and inventory control type applications. As this technology becomes used on a larger scale it is expected the tags will be available for (USD) $0.10

each or less. However they are limited in the fact that they can only provide location information in conjunction with a software application.

5.1.1 Description

We developed a simple application as an introduction to the hardware components and how to collaborate between software applications and RFID tags. The purpose of the application is to provide a management system which can locate assets and record their current location and keep a record their history. This information can then be displayed to the user through a software application (software flow shown in Figure 3).

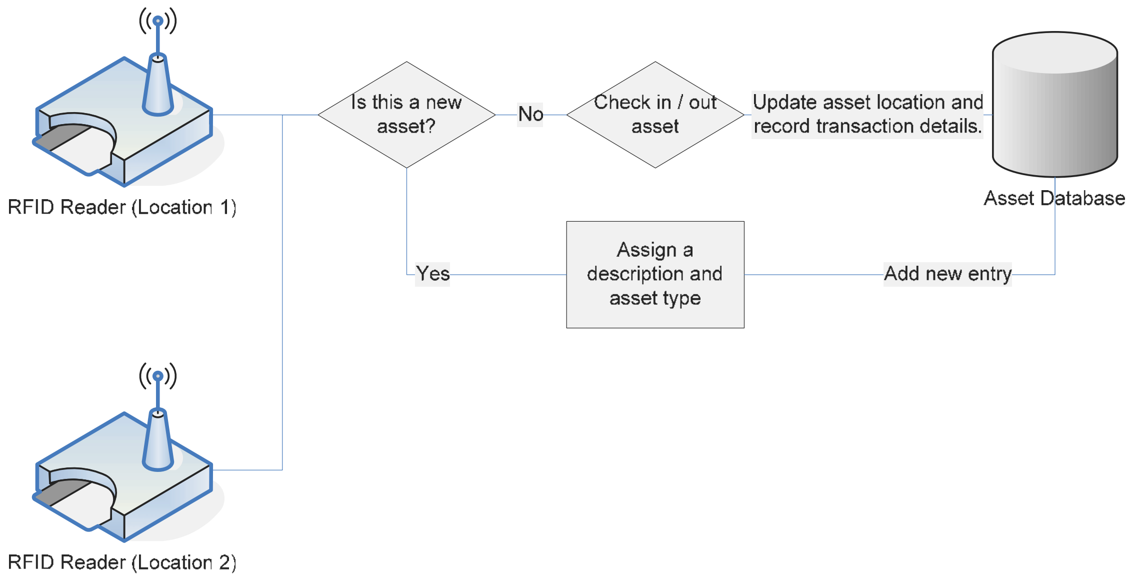


Figure 3: Software flow for Configuration Management

Tags were attached to a box (Figure 5), and the box was assumed as an asset, with the contents of box as the configuration or the constituents of an asset. Our aim is to monitor the location of each configuration item as well as that of the asset. In the application that we developed, when an asset is first read at a location the user is asked to enter a description and define the type of asset. There are two types of assets we have used in our design, items (configuration items) and containers (assets). An item is the lowest level of the two and only has knowledge of its own existence. Containers on the other hand can hold items within themselves. When an item type is selected this asset is added to the database and is now ready normal use. Now as an asset leaves and enters a room it is read by the reader which in turn adds transactions to the database. Each transaction records the location of the tag and the time it occurred. When a container type is selected the operation is the similar to that of an item with the exception an item may be placed within a container (Figure 4). This allows us to keep track of groups of related objects as one entity or asses when querying the database.

Figure 4: Inlay RFID Tag attached to container using a stick on plastic envelope.

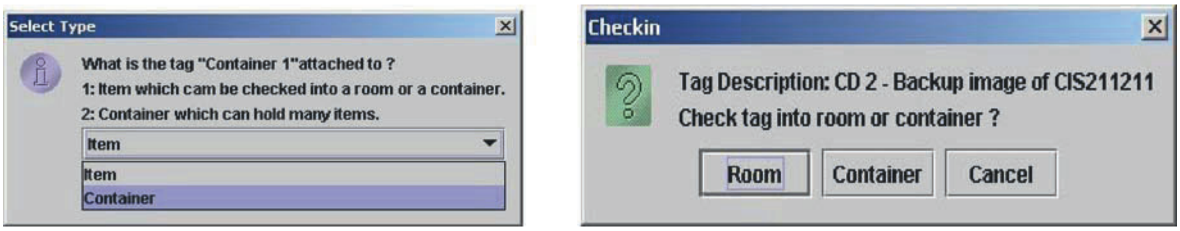


Figure 5

Figure 6 illustrates the snapshot of the user interface that describes the containers and the items contained in there.

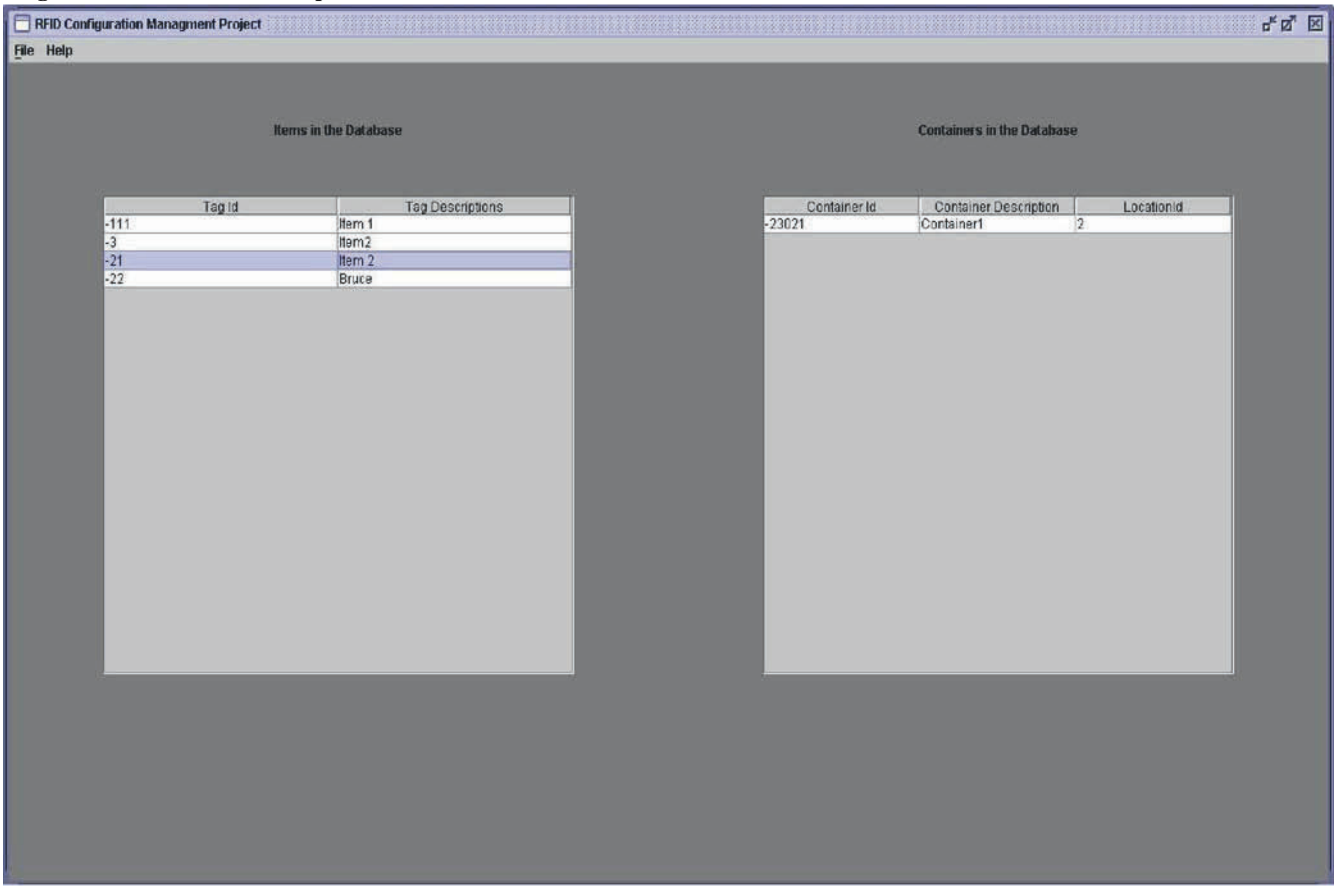


Figure 6: Demo Application Snapshot

One of the issues in configuration management is that of change management. Therefore it was necessary to capture the history of configuration. An example of each of this query is shown in figure 7.

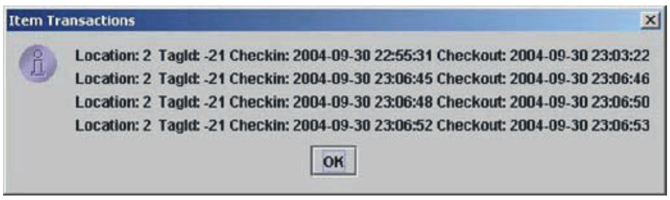


Figure 7: Configuration item history details

5.1.2 Issues

One of the limitations which should be considered is the material in which tags are attached to as well as the materials in the environment. When the tags are attached to metallic surface the tag malfunctions. Texas Instruments have released a range of tags designed to be mounted on metallic surfaces. However their reading range is reduced by 30% and their size is much larger than the inlay variety.

5.1.3 Future Directions

It appears that in highly metallic assets such as railway tracks, rolling stock, or military hardware existing RFID tags, especially the inlays, may not function. It is therefore, necessary to investigate how a robust tag can be developed that is able to provide the same functionality in a variety of environments. In addition to this, the issues regarding RFID technology working on different frequency spectrum will remain, particularly when assets are moved between international borders.

5.2 RFID for Supply Chain Management

Using the same methodology, as described above, we examined the Texas Instruments Series 6000 - HF-I Midrange technology. The midrange evaluation kit consists of one reader and a selection of tags in different packages. It is a high frequency kit, which operates at 13.56MHz. The reader is capable of reading tags at a range of 7 inches and its plastic housing integrates an antenna and an RS232 interface board. The reader is able to communicate with Tag-It as well as any ISO 15693 compatible tags. We used two readers in this scenario, with the other reader being S251B, which is a low frequency reader operating at 134.2 kHz and is able to communicate with TIRIS Low Frequency tags. It automatically tunes a standard antenna to resonance and keeps it tuned during operation. It is able to communicate with tags at ranges up to 1m (using a large size gate antenna). The reader decodes the signals emitted from transponders into identification numbers, performs error checking and translates the signals to standard serial interface protocols (RS232, RS422/485). It also contains a buffer, which can store up to 909 reading transactions. Communication with a PC can be performed over a serial interface (RS232) using ASCII commands.

5.2.1 Prototype Application

The prototype asset inventory management application developed using the Series 6000 - HF-I Midrange technology, mimics the scenario of movement of assets or items between a field forces distribution centre (FDC) and field supply depot (FSD). The application displays two warehouses i.e. of FDC and FSD, each containing six sections. If each section was equipped with a reader, we could easily locate the section containing an item at any time. The two readers are installed at the entrances to the warehouses, thereby allowing for tracking movement of tagged items, Figure 11.

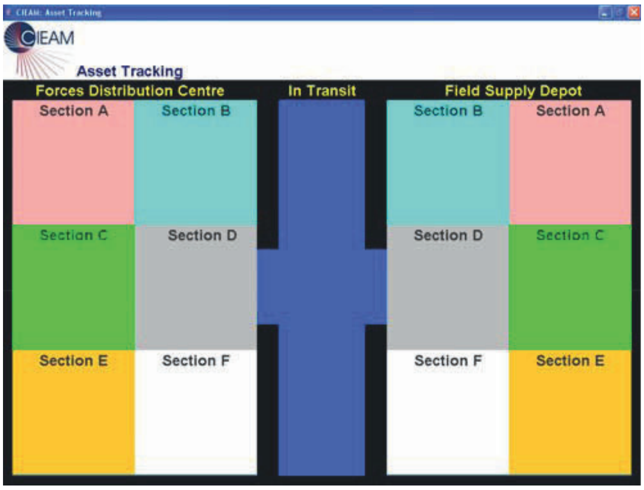


Figure 8: Asset Tracking Application Interface

When an item enters Forces Distribution Centre for the first time, the user is asked to enter item details figure 9. After entering item details, the item is displayed in the system, shown in figure 10.

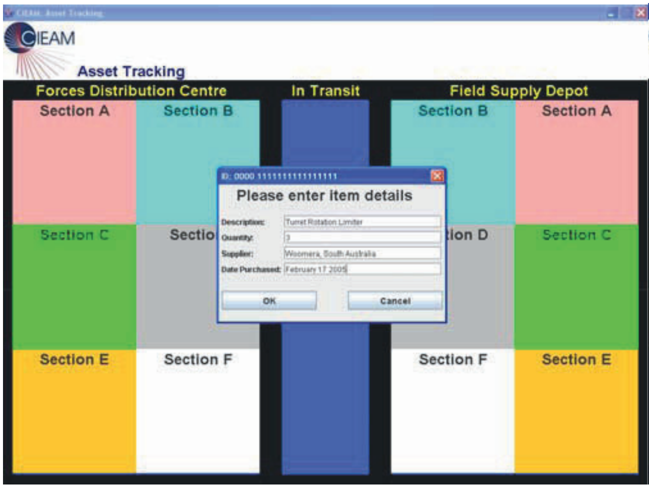


Figure 9:

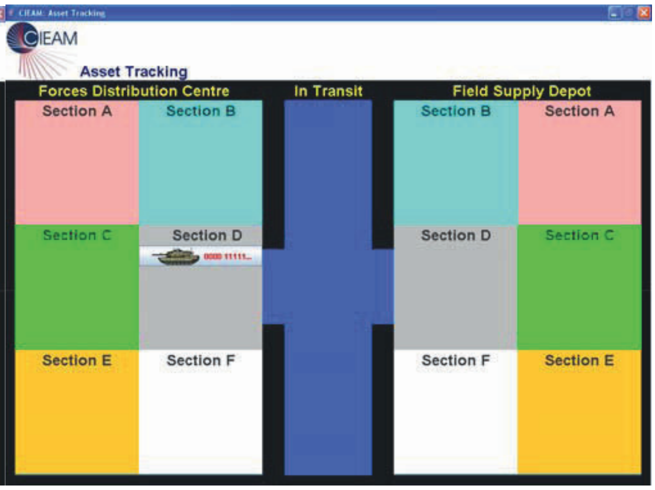


Figure 10:

When an item is moved out of the warehouse, it displays the status of the item as in transit, till the time it is checked in another warehouse. (Figure11). Figure 12 illustrates that history of an item's movement along with the time and date of each movement.

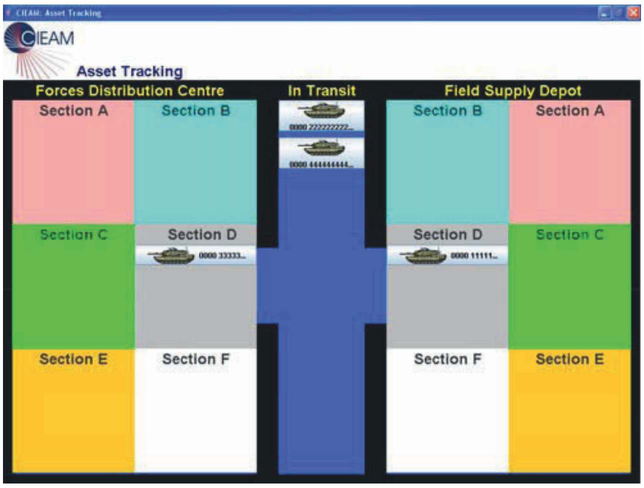


Figure 11:

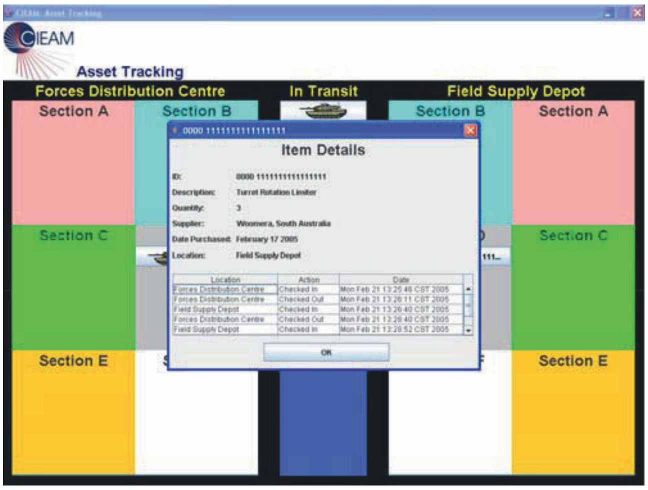


Figure 12:

However, it must be pointed out that the prototype was tested in laboratory only and did not have the chance to test it in field. Nevertheless, some of the issues that we faced using this technology were the malfunctioning of the technology in high magnetic fields, and in environments where other high frequencies based equipment was in operation.

5.2.2 Future Directions

Apart from the issues identified for configuration management, a major issue that we encountered for item tracking was the real time monitoring of an item when the same is in transit. RFID tags have a limitation and require the reader to be in close proximity of a tag, which means that they work well in the warehouse or a defined space, however cannot be used to track items moving out of the range. We considered GPS tags; however GPS technology cannot work in in-door environment. Therefore, in its existing state RFID technology can be used for item tracking between different staging sections and within those staging section; however it cannot be tracked in real-time when the item moves out of the staging sections. Further research is required in resolving this issue.

5.3 RFID for Condition Monitoring

Chipcon's CC1010 evaluation kit includes an evaluation board (CC1010EB) and two evaluation modules (CC1010EM). CC1010EB provides access to all of the analogue and digital pins on the CC1010, along with two serial ports, a parallel programming port, RF network analysis ports, and other peripherals. Each evaluation module features CC1010, an antenna port, and an analogue temperature sensor. CC1010EM includes an RF transceiver that can run from 300 MHz through to 1GHz. It has an onboard 8051 8-bit microcontroller with 32Kb of flash for code and non-volatile data, 2K of RAM. CC1010EB supports three ADC (analogue-to-digital converter) channels, a Universal Asynchronous Receiver Transmitter (UART), and 26 I/O pins. An important feature of CC1010 is that it can be read and written from distances in excess of 100 meters. At the same time, the tag supports a serial peripheral interface, and a hardware Data Encryption Standard chip for secure communication. Its flash RAM is divided into 256 pages with programmable protection flags that prevent unauthorized downloading of internal programs and data. All these features make the CC1010 a versatile platform for applications where we want to add external sensors to the modules. The general purpose I/O pins can be connecter to temperature, humidity, pressure and a whole range of external sensors. The two hardware UARTS are used to communicate between serial devices, this may be a PC other RFID modules or and hardware device using serial communications.

Figure 13: Left CC1010 module, Right CC1010 development board with module attached used for debugging purposes.

We envisioned the framework presented in figure 14 that utilizes CC1010 tags. In this framework each asset is attached with a CC1010 tag or a collection of it, such that all tags collectively form a wireless mote network with the provision of a gateway tag that provides for communication of data captured by the network. The gateway tag serves as the liaison between the database server and sensor network and delivers data captured to condition monitoring database. Apart from unique identification of the asset, the availability of serial ports in CC1010 provides for interfacing digital as well as analogue sensors. This means that it allows a connection with the any amount of sensors used for condition monitoring of an asset. This framework highlights continuous streaming of data without having to have any human intervention for data logging or recording. Obvious advantages of this type of CM are reduced labour cost, better fault detection, and identification of dangerous or hazardous areas without incurring risks to maintenance crew. Most significant advantage of this framework is that it makes possible economical remote condition monitoring. The framework captures data from different channels simultaneously and processing the same immediately, thereby by achieving greater accuracy in failure prediction. At the same time, it is also useful in isolating the area where the fault may be developing.

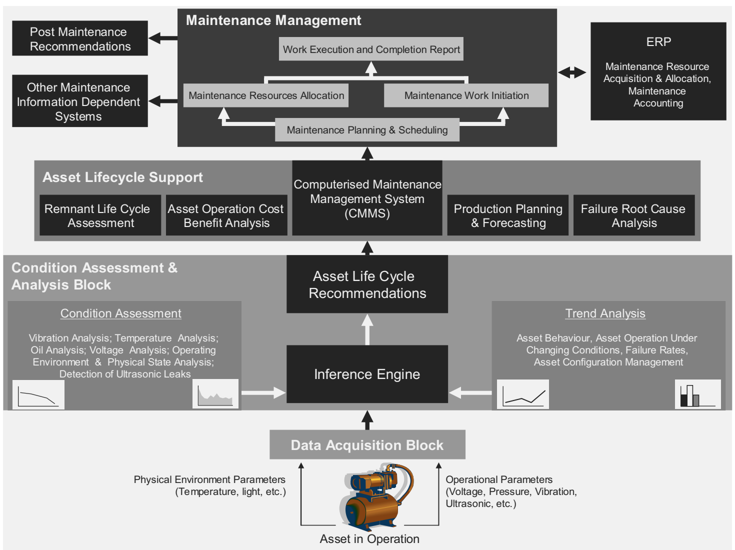


Figure 14: Integrated Condition Assessment Framework

In this proposed framework, the condition assessment block represents a series of decision support indicators, analyses and trend projections of assets operation. Using these comparisons and behavior assessments of asset operation useful deductions could be drawn for asset lifecycle support. The inference engine acts like a watch dog and compares the probabilities and possibilities before generating recommendations concerning asset life cycle support and other associated processes. Using RFID for prognostics poses significant advantages for engineering enterprises, and proves to be an important building block towards e-intelligence based maintenance and asset management. This framework allows for a technologically convergent and miniaturized device that bears most of the features of a SCADA systems at a fraction of its price.

For proof of concept we used the technologies presented in figure 15, and an external humidity sensor and the temperature sensor that is embedded in CC1010. The raw information captured from the sensors was communicated through the gateway to a transition database. The interface allowed for viewing this information as graphs or simple readings over a period of time. However, this information could easily be fed into a condition database with time stamp and tag ID, where it could be analysed further. The TCP/IP based connections allowed for a distributed server model, which means that using this framework remote condition monitoring is also possible. Apart from providing advances such as consolidation of condition information and single channel communication, another advantage of this framework is fault detection. An essential characteristic of RFID tag is ability to identify itself, therefore when the information captured from each asset is analysed, it provides important indication on exact location of the asset under investigation. For example in case of a failure condition developing or any malfunctioning it can point that at site number 1, pump no 2's vibration pattern is unusual. At the same time, this information together with information on other variables will provide for the exact nature of the fault and indications on its maintenance demands. The onboard memory provision of CC1010 could be used to store information, such as asset health and maintenance record, which could be updated after each maintenance execution.

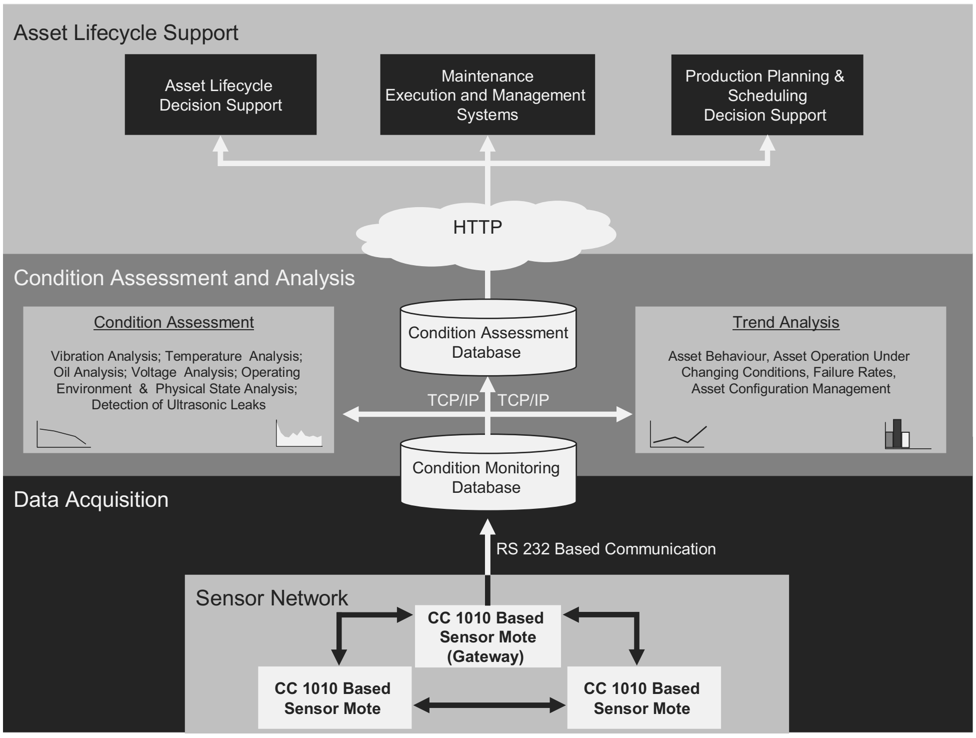


Figure 15: Integrated Condition Assessment with CC1010

Using RFID for prognostics poses significant advantages for engineering enterprises, and may prove to be an important building block towards e-intelligence based maintenance and asset management. This methodology allows for development of a technologically convergent and miniaturized device that could easily be mounted on an asset and realise centralized data acquisition. It reduces monitoring costs significantly and improves asset reliability through efficient prediction of asset failures.

5.3.1 Future Directions

We tested the technology for two sensors; it will be interesting to see how the RFID tag behaves when more sensors are attached. Other area of research is of making the technology more efficient, as the on board power supply of CC1010 has limited life span. Once these issues are resolved, it will be worth while to investigate the ways of integrating the inference engine on the gateway sensor mote.

6 CHALLENGES POSED TO RFID

Since the modern wave of RFID technology is still evolving and the technology can best be described as emergent, there are many issues and challenges associated with the technology. As mentioned before, one of the issues associated with RFID is the lack of uniform frequency spectrum. Different standards organisations and governmental legislations are under way to address this issue. Standards development efforts also include developing standards for hardware and software used in RFID technologies. However, there is another issue that emerges from standards development, which is regarding the multiple efforts towards developing standards for RFID operation. This means that in the presence of multiple standards there is no guarantee that the technology will embrace uniformity at a global scale. A corollary to the issue of multiplicity of standards is interoperability. This is, however, not to say that this problem does not exist at the moment. The issue of interoperability includes readers as well as tags, as presently the tags and readers have fixed operations and it is not possible to read high frequency tags with a low frequency reader. Acceptance of RFID technology will require it to be of low cost, and to be of low cost this issue of interoperability needs to be resolved. Cost effectiveness of RFID technology is actually an issue that small business and SMEs are facing at the moment, as they have limited budgets and limited expertise available at their disposal. Security and privacy are other two issues that are catching the attention of consumers as well as general public worldwide. With the existing nature of technology (primarily passive RFID tags) it is possible to read tags without any problem, which means that anybody could identify tagged items and could track their movement. A solution often proposed to address this issue is to encrypt the message, however as yet there is no solution on market that could provide this functionality and security.

7 CONCLUSION AND RECOMMENDATIONS

The primary aim of this evaluation study was to test RFID technology for configuration management; however we were able to successfully test and demonstrate the same for three different applications. During the course we also uncovered the potential of RFID technologies with regards to asset maintenance. Nevertheless, the proliferation of RFID technology is subject to the development of technology to overcome some of the issues identified in the paper. These issues and challenges include robustness of technology, operation of technology in different environments, lack of standards, cost, data/infrastructure/systems implementation issues, data sharing and security. Nevertheless, with further research into these areas it is anticipated that large scale implementation of RFID will occur in the next five years.

8 REFERENCE

[1] d'Hont, S. International Newsletter of the TI RFID, accessed online at http://www.ti.com/tiris/docs/manuals/RFIDNews/Tiris_NL20.pdf , on December 30, 2005.

[2] AutoID (2004) Technology Guide. *AutoID Centre* online accessed at http://quintessenz.org/rfid-docs/www.autoidcenter.org/new_media/brochures/Technology_Guide.pdf, on January 10, 2006

[3] Haller H & Hodges S (2002) White paper: The Need for a Universal Smart Sensor Network, *Auto-ID centre Institute for Manufacturing*, University of Cambridge, United Kingdom, accessed online at http://www-mmd.eng.cam.ac.uk/automation/w_papers/cam-autoid-wh012.pdf , on December 30, 2005.

[4] Chipcon SmartRF ® CC1010 PRELIMINARY Datasheet (rev. 1.2), Chipcon, accessed online at http://www.chipcon.com/files/CC1010_Data_Sheet_1_2.pdf , on December 30, 2005.2005.

[5] NIST(2004), "RFID explained: A Basic Overview", National Institute of Standards and Technology, February.

[6] NIST(2004), "Supply Chain Technology", National Institute of Standards and Technology, January.

[7] Floerkemeier C Lampe M & Schoch T (2003) The Smart Box Concept for Ubiquitous Computing Environments, *Proceedings of Smart Objects Conference*, Grenoble.

[8] Levinson M (2003) All-in-One Appliance - THE REFRIGERATOR, *CIO Magazine*, February.

[9] Wan D (2000) MAGIC WARDROBE: Situated Shopping from Your Own Bedroom, *Proceedings of the Second International Symposium on Handheld and Ubiquitous Computing*, Bristol, UK.

[10] Bubny P (2000) Keeping cool with RFIDs. *Supermarket Business*, 55(5), 95.

[11] Raman A DeHoratius N & Zeynep T (2001) Execution: the missing link in retail operations, *California Management Review*, 43(3), 136-52.

[12] GAO, (2005), "Information Security : Radio Frequency Identification Technology in the Federal Government", Report Number GAO-05-551, United States Government Accountability Office, Washington, D.C. 20548

Acknowledgments

This paper is a part of the research conducted for CRC Integrated engineering Assets Management (CIEAM), at the University of South Australia. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the CIEAM.

INTELLIGENT MAINTENANCE SYSTEM APPLICATION TO GLUE DISPENSING MACHINE

J. W. Chen, J. N. Jeng, Z. C. Chou, T. K. Yang and H. H. Li

RTC/ITRI and MSL/ITRI, Industrial Technology Research Institute (Taiwan)

Abstract: The tiny hole at the needle distal of the LCD (Liquid Crystal Display) Dispenser is frequently blocked due to the impure glue or clotted glue. Although many research studies have been performed for Intelligent Maintenance System (IMS), few of them are relevant to the discussion on the result of the process in this respect. For the most part, the downtime of a packaging production line occurs in the glue dispensing process. This paper is related to the feasibility study for the IMS application on the glue dispensing machine. It not only possesses the capability of self-diagnosis and self-improvement by employing effective data learning, but also ensures the equipment works safely for a long time. To implement such a diagnosis system, the raw data is collected directly from the images with the cameras on the equipment, and these data provided for the Back-Propagation Neural Network system to learn. The Back-Propagation Neural Network system is applicable to predicate these kinds of failure problems. The feasibility study of forming a knowledge exploration and intelligent decision-making system by making use of the theory of the Back-Propagation Neural Network system, it can be discussed through a case study.

Keywords: Glue Dispensing, Glue Path, Back-Propagation Neural Network, Nearest Neighbor Decision Rule, Vision Assistant

1 INTRODUCTION

The automatic industry fluids process is often intricate and expensive. Its applicative scope includes fluids to strong creams. These kinds of materials need to handle carefully and to reach the quality of consistency in the assembling process [1]. In order to make perfect for spreading fluid glue, it can control by atmospheric pressure or mechanical power device or desktop automatic Dispenser equipment, its application is very extensive. In generally, the state of syringe needle depend on the quality of glue, with spreading fluid glue in succession will bring some glue on the needle become the dregs of glue, then it is blocked the needle distal as shown in Fig. 1.

Fig. 1. Glue syringe needles with incomplete glue

The hardened dregs of glue will influence gradually the state of glued normally, so it is apt to trouble phenomenon of syringe needle blocking, skewing or out of shape, etc. The caused of failure phenomenon produced is even probably a bad needles distal. It is not to conform to the automatic and production line increasingly required, about deciding when they need to change the glue needles distal by human eye, and it is unable to control question which is an individual judgment subjectively about the glue quality determining with operator's experience of glue states determining when the suitable for glue needles distal change. In order to improve the quality of dispensing process, and improving the competitiveness for the products in the market, this paper is focal point on comprehensive quality of glued determining.

Recently the machine vision components are using for product inspecting include image acquisition card, CCD camera, CCD lens, and auxiliary light source, etc. Artificial Neural Network has already become one of the important researches of artificial intelligence (AI) themes in the field in recent years. The main reason is the Artificial Neural Network is possessed of learning ability, memory ability, summing up ability, fault-tolerant ability, and high-speed computing capability, etc [2, 3].

In order to determine the occasion for changing glue needle distal, this research use Back-Propagation Network and nearest neighbor decision rule to combine the vision system and measure a flaw glue to achieve the automatic vision and measure to replace the human eye estimated.

2 STRUCTURE AND PROCESS

The purpose of this paper employs Back-Propagation Network and nearest neighbor decision rules to measures flaw glue for determine the occasion to change glue needle distal. This structure includes four parts: system equipment, image process, Back-Propagation Neural Network and nearest neighbor decision rule.

Fig. 2. Dispensing machine platforms

The system hardware configurations are included dispensing machine platforms of desktop on PC-based that production line environment, CCD camera, auxiliary light source and image acquisition card as shown in Fig. 2. Image process is used by NI Vision assistant software. The CCD camera is set up on a column base beside the glue needle; through the glue of glue needle process, the camera can successively take a picture passing on glue dispensing path. Then utilize NI Vision Assistant to deal with the picture of feature. Because there is no real working environment of production line, therefore an experiment goes on by way of simulate a real factory and working for a long time. In order to reach quickly the situation about glue needle distal of failure, the experiment adopts one gluing process to finish and then shut down experiment for one hour later; that will accelerates the glue of needle distal clotted. This process is repeating until glue needle distal is clotted, finally the tiny hole at the needle distal of the dispenser is blocked by the glue clot and unable to glue.

Adjusting auxiliary light source and contrast for making image edge obviously while taking a photograph. The photograph taken will be dealt with the tool of package software of Vision assistant and subsequently threshold process to separate foreground from background. Therefore, to employ Low Pass Filter to eliminate noise mixed point and unnecessary border. These pictures were photographed by camera can regarded as input patterns for recognized. Use the conventional back-propagation model to train those patterns that have previous took photographs by camera. As a result, the search model can respond quickly the desired results whenever similar photograph have applied to the dispenser. Not only the designing procedures have been discussed but also the simulation results are presented to validate the effectiveness of this research. There are many phenomenon of glue needle blocking and its photo patterns include normal, skew, thin, width and out of shaping type, etc. The result of back-propagation model inferred will be given nearest neighbor decision rule to decision whether change glue syringe needle.

The nearest neighbor decision rule is a commonly used for classification algorithm in statistical pattern recognition. Each class is given as a set of sample prototypes, a training set of pattern vectors from that class. When an unknown vector is to be classified, its closest neighbors are found from among all the prototype vectors, and the class label is decided based on a majority rule. This rule is simple and elegant, and yet the error rate is small in practice. The block diagram of the Intelligent Maintenance System for Glue Dispensing Machine is depicted in Fig. 3.

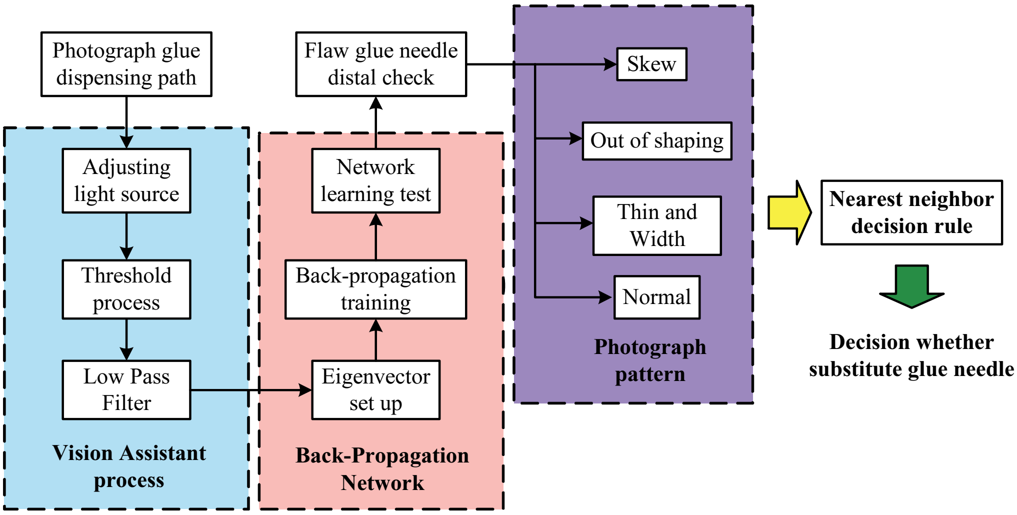


Fig. 3. Block diagram of intelligent maintenance system for glue dispensing machine

3 THE IMAGE PROCESS FOR CHECK FLAW IN GLUE PATH

Fig. 4. The result of threshold process

In practical application of image process, if the main object can be deal with a background which is not necessary to separate, the later process will more easily. It is easier to design Back-Propagation Network algorithms on binary images than on gray-level images. A binary image can be obtained by threshold a gray-level image. For preventing possible information losses in the threshold process, it may be natural to design Back-Propagation Network algorithms directly on the original gray-level images. The gray-level image is a binary image constructed by bright background and dark object. Thus, the bright background and gray main object have been divided into two parts. The best way to extract main object from background image is choosing one critical value T that can obviously separate it into parts. In this paper, the critical value T is 80 and user decides that. A binary image can be defined as g(x, y):

$$g(x,y) = \begin{cases} 1 & f(x,y) > T \\ 0 & f(x,y) \le T \end{cases}$$

These marks are correspondent to the background for the picture pixel of 1 and the other mark is correspondent to objects for the picture pixel of 0. The result of threshold process of image glue path is depicted in Fig. 4. It is clear to discern when the value T be set up suitable, the glue path will be separated obviously from background.

Fig. 5. With noise picture

The photograph noise means that the image structure to the true scenery does not exist in fact. The noise will influence the judgment on the true signal. For example in the one-dimensional image is shown in Fig. 5., nearby signal on the left of the glue path is fuzzy, that is to say that the noise is very strong, we will infer that the authenticity of the signal in pictures, perhaps it is only protruding higher noise of a coincidence. In Fig. 5., through the threshold process is unable to strengthen this signal directly because the raw data signal with a low credibility and noise will be enhanced at the same time. In Fig. 6, after filter process, edge value comparatively obvious, and treatment in follow-up will be more convenient comparatively and result of value can comparatively perfect, too. It can reduce the situation of judging by accident.

Fig. 6. After noise filter process

The noise influence image process that limits the image to deal with deepness. It is impossible for this restriction to disappear. However, it can enhance to the image process by some methods then the enlarged noise image is not obviously to keep the image intact. This method is regarded as reduce the tiny noise of structure in the image which is not significant at the cost of losing precision order of image [4].

Fig. 7. The raw data of image picture

4 EXPERIMENT AND RESULT

In this research, it is using the Vision Assistant 7.1 of National Instruments of package software [5], image acquisition (IMAQ PCI-1405) device, SONY XC-ES50 (CCD) camera, 1x65 extending camera lenses, and STDR 67/36R auxiliary light source. In the course of experiment, utilize CCD lens and extending lens to take an image picture that is glue path as shown in Fig. 7. After storing the image in the computer, use Vision assistant as the follow-up to deal with image process and analyze.

Open the image by Vision assistant, using Brightness option adjusts the brightness and contrast of image as shown in Fig. 8. This is easily to make the edge of picture obviously and facilitates the follow-up image process. The gray-level image is threshold processed by Gray Morphology, as shown in Fig. 4. The main objects will be separated with background which is not needed existing. Choosing a critical value T = 80 that can obviously separate it into parts and makes the mark correspondent to the background for the picture pixel of 1 and other marks are correspondent to main objects for the picture pixel of 0. When the critical value T is proper set up, the glue path will be separated from background clearly. In order to provide better picture quality and measure glue path easily, it must to reduce image noise and unnecessary edge. In this paper, it uses FFT Filter function of Vision Assistant Package software to improve the middle weighting value, as shown in Figs. 5, 6. The filters function as:

$$\begin{bmatrix} 1/10 & 1/10 & 1/10 \\ 1/10 & 2/10 & 1/10 \\ 1/10 & 1/10 & 1/10 \end{bmatrix}$$

4.1 Pattern parameter

Picking image of image process module makes the standard pattern parameter and flaw pattern parameter to treat as studying the module group in the pattern. This is the eigenvector needed to use that training study for the setting up type neural network. The pattern parameter mode is:

(1) Input the eigenvector of the parameter:

a. $\dfrac{\text{width of glue path}}{\text{width of picture}}$, determine whether the glue path is too width or thin.

b. $\dfrac{\text{skew of glue path}}{(\text{average width of glue path})^{1/2}}$, determine the skew of the glue path.

c. |prior slope of invert – posterior slope of invert|, determine the out of shaping of the glue of path.

(2) The eigenvalue of output parameters are 6 determined results as follows by:

a. left skew
b. right skew
c. left out of shaping

d. right out of shaping
e. thin
f. width

There are 13 picked points of the type parameters of glue path, which one of them is as shown in table 1. Table 1 is the normal case of glue path pattern and others case are omitted from here. The way of picked type parameters is:

A: $\dfrac{\text{width of glue path}}{\text{width of picture}}$, B: $\dfrac{\text{skew of glue path}}{(\text{average width of glue path})^{1/2}}$

C: left side |prior slope of invert – posterior slope of invert|

D: right side |prior slope of invert – posterior slope of invert|

Table 1. Normal glue path pattern (Unit: Pixel)

Normal glue path	position		A	B	C	D
	X	Y				
1	144.3845	36	0.5235	-0.1978	0.0301	0.0262
	479.4232	36				
2	144.2936	56	0.5224	-0.2087	0.0008	0.0161
	478.6279	56				
3	144.1335	86	0.5215	-0.2193	0.0212	0.0120
	477.9194	86				
4	143.4707	111	0.5221	-0.2310	0.0128	0.0086
	477.6279	111				
5	143.031	143	0.5227	-0.2375	0.0037	0.0214
	477.5293	143				
6	142.4382	177	0.5223	-0.2549	0.0080	0.0175
	476.697	177				
7	142.126	210	0.5224	-0.2616	0.0145	0.0163
	476.4673	210				
8	141.2378	247	0.5225	-0.2829	0.0019	0.0193
	475.6069	247				

9	140.4351	278	0.5235	-0.2942	0.0006	0.0117
	475.4849	278				
10	139.4824	314	0.5241	-0.3127	0.0211	0.0061
	474.9219	314				
11	139.3149	345	0.5239	-0.3184	0.0222	0.0167
	474.627	345				
12	138.5151	374	0.5240	-0.3374	0.0177	0.0228
	473.8682	374				
13	138.2686	399	0.5242	-0.3415	0.0203	0.0022
	473.7842	399				

4.2 Back-Propagation Network Learning

After getting parameters of the glue path types, that then going on the Back-Propagation Network learning, which will recall the network weighting value, needed to use while inferring by receiving the network. The measuring project of this system is divided into four types, such as normal, width, thin, skew of glue path and out of shaping, etc. Therefore, building the network of constructional learning which necessary learning parameters is every type of glue path to pick up 5 pictures and a picture is divided apart into 17 partitions that are 15 lines and using among of 13 lines have 26 points, as depicted in Fig. 8. Since there are 6 eigenvalues of output parameters, the total learning patterns are 390. Because of it is too huge to show training data of the parameters, there are 13 data among them as shown in table 2.

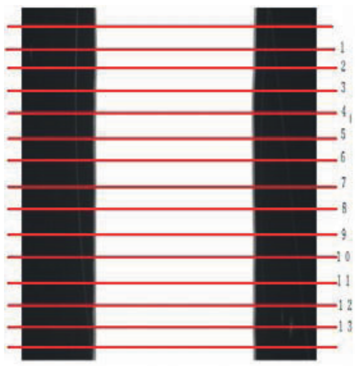


Fig. 8. Normal pattern

Table 2 Training data

	A	B	C	D	left skew	right skew	left out of shaping	right out of shaping	thin	width
1	0.5235	-0.1978	0.0301	0.0262	0	0	0	0	0	0
2	0.5224	-0.2087	0.0008	0.0161	0	0	0	0	0	0
3	0.5215	-0.2193	0.0212	0.0120	0	0	0	0	0	0
4	0.5221	-0.2310	0.0128	0.0086	0	0	0	0	0	0
5	0.5227	-0.2375	0.0037	0.0214	0	0	0	0	0	0

6	0.5223	-0.2549	0.0080	0.0175	0	0	0	0	0	0
7	0.5224	-0.2616	0.0145	0.0163	0	0	0	0	0	0
8	0.5225	-0.2829	0.0019	0.0193	0	0	0	0	0	0
9	0.5235	-0.2942	0.0006	0.0117	0	0	0	0	0	0
10	0.5241	-0.3127	0.0211	0.0061	0	0	0	0	0	0
11	0.5239	-0.3184	0.0222	0.0167	0	0	0	0	0	0
12	0.5240	-0.3374	0.0177	0.0228	0	0	0	0	0	0
13	0.5242	-0.3415	0.0203	0.0022	0	0	0	0	0	0

The constructional learning data can input Back-Propagation Network learning module to learn and then obtain a network link weighting value that support follow-up spend as what has been inferred to come. In this study, the Back-Propagation Network adopt input layer, two hidden layers and output layer altogether 4 structures and their states set up as follows:

1. Input layer nodes are the input parameters that is the mentioned 3 eigenvalues previously.
2. Output layer nodes are the network-outputted layer to output several nodes that is the 6 kinds of glue path types mentioned previously.
3. There are 390 training patterns to learn in the network. The more network train epochs, the more convergences disappear. In the experiment, training epoch is setting 100 times and disappears of that exceed the number of times, but increasing the time to train instead.
4. The hidden nodes are the process numbers of hidden layer, too much or too few affect the convergence result. The result of the hidden layer cannot reflect this problem and input the reciprocation among their units. Therefore, there are more errors. One or two hidden layers can overcome this more complex problem and reciprocate among their units of the response. Too more hidden layers make the network excessively complicated instead and more some minimums that will make also falls into some minimum of an error function even more while revising the weighting value of the network, it is unable to convergence. According to experience, the hidden layer of general problem is one layer and the more complicated problem is two layers. In this system, we take two hidden layers and the nodes of two layers are 10 respectively. The error convergence figure of after the network finishes learning is depicted in Fig. 9.

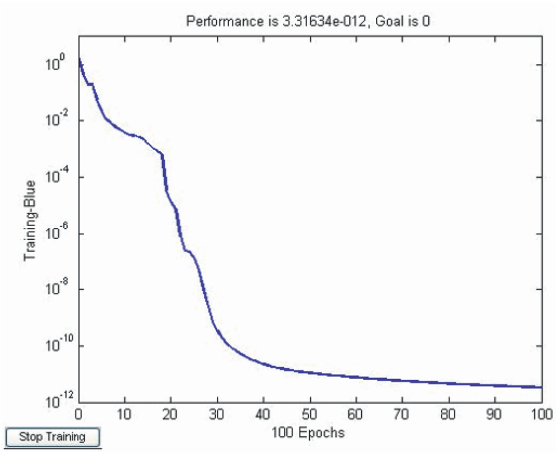


Fig. 9. The error of network learning

After 100 epochs of training, the root mean square error is converged to below 3.316×10^{-12} and the result of training network weighting value is as follows table 3 and table 4. The result of example training test is as shown in Fig. 10.

Table 3. First layer weighting values

-9.2937	-1.2963	-0.0268	6.0602
-4.5012	-0.7557	-0.0676	7.8189
-0.1778	6.2812	0.6369	-0.8289
19.0420	0.2600	2.1741	-25.8086
13.8294	1.7206	-10.9858	-1.7883
-5.2034	0.0412	-1.7951	-27.2137
-5.0449	-0.4220	4.5244	13.5043
6.9237	-0.0005	16.0749	1.1656
0.8282	0.3130	-2.7430	-5.5210
-3.2390	3.6022	8.7540	-0.8492

Table 4. Second layer weighting values

10.1959	3.6028	17.006	-14.0076	-0.7204	-10.4126	9.7959	46.7295	-5.5545	-0.9451
-4.5026	5.2559	-16.1377	15.5656	1.4934	1.9922	-7.3308	0.3930	4.8295	-7.0968
-63.7379	-30.5908	5.4719	-85.4169	-11.7628	30.1749	-0.5691	-0.4885	-7.2311	-0.1325
-12.8968	55.8255	22.2025	38.4423	12.6196	-13.1873	-40.3262	-16.7146	-9.3594	5.0204
-12.0259	0.3903	0.8283	11.1641	1.4007	-18.0422	6.3489	-1.2162	5.9670	12.1222
-4.5530	5.7189	-15.9568	15.4884	19.7044	-10.7731	-11.0754	0.3985	2.2986	-7.2016
2.1680	6.0024	21.9208	-21.7898	-11.8160	-9.2856	3.3616	0.9328	0.7776	0.3456
3.2484	-0.9094	-0.5516	-1.5583	-0.0829	0.4089	-3.6165	0.0336	6.1276	2.3997
-5.2448	1.7923	-2.1697	1.0190	-3.4291	-17.2132	-8.2941	1.0343	-2.3586	-6.2704
-20.7555	-55.5086	-8.4918	-38.4608	2.4239	27.6850	-8.9109	16.5992	-8.6939	-14.5556

4.3 Glue path check

The final neural network learning is checking process. According to these link-weighting values of networks, it determines the output type of that will be measured event. This approaches to determine the failure type of glue path and replace the glue needle decision. The following each measure sample, the result of regular output is between 0 and 1. This is significant of it is closer to 1 and then the features of classified are more obvious. On the contrary, it is closer to 0 and the features of classified aren't obvious. According to experiment, while the network infers for some types that exporting values are very close to 1 and others are 0.

4.4 Nearest neighbour decision rule

The nearest neighbour decision rule itself has not included too many foundations of theorem [6]. It depends on intuition way to decide that how to classify who is the neighbour and who is closest to decide one's own classification. If there is a vector v of sample of an unknown classification, then we find out the closest point v_k to v in all sample data and use v_k to decide the v to classify which category. It is satisfied as following conditions that v and v_k can include in the same category and the mathematics form can express to:

$$dist(v-v_k) \leq dist(v-v_i),\ i = 1 \ldots n,\ i \neq k$$

Where $dist(v-v_k)$ is any way to distance measure, for example Euclidean distance.

The case of Fig. 11 is the Prototype of test pattern and its result of Back-Propagation Network training is as shown in table 5. The normal state position is defined x=0 and others failure state skew, out of shaping and thin-width are defined x=1, x=2 and x=3 respectively. Fig. 12 is shown 13 sample points distribution state. It indicate that the sum of Euclidean distance It calculate sum of Euclidean distance between the failure position and 13 sample points then it can find that the position skew x=1 is the sum of minimum distance. Therefore it can infer that the most possibly failure type is a skew state and advised to change glue needle distal.

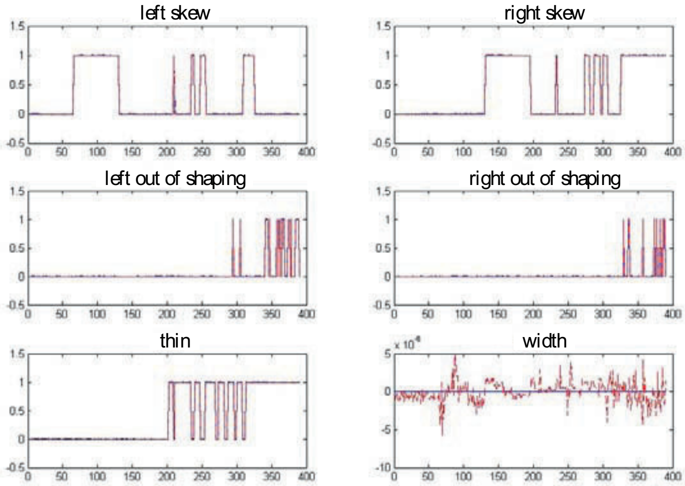


Fig. 10. The result of training test

Table 5. Result of pattern Back-Propagation Network training

	1	2	3	4	5	6	7	8	9	10	11	12	13
Skew	1	1	1	1	1	1	1	1	1	1	0	0	0
out of shaping	0	0	0	0	0	0	0	0	0	0	0	0	0
thin and width	0	0	0	0	1	1	1	1	0	0	0	0	0

Fig. 11. Prototype of test pattern

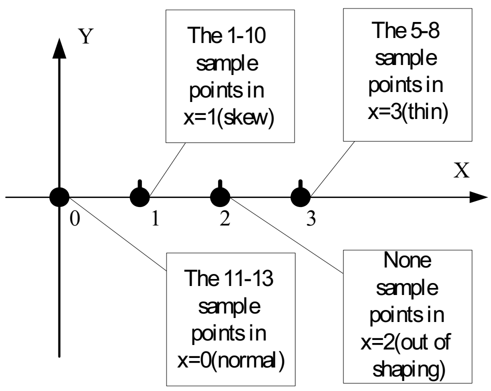


Fig. 12. Distribution of sample point

5 CONCLUSION

In this research, it is using Back-Propagation Network, nearest neighbor decision rule and Vision assistant system to develop an intelligent equipment maintenance system for the diagnostic research. The main diagnostics is to diagnose with the automatic vision process that can reduce the careless mistake from human eye and subjective experience. According to much of the sample test, this diagnosis system is base on Back-Propagation Neural Network has already a precise distinguished. In the future, more reference data can even enhance the height ability.

6 6. REFERENCE

[1] FALCON AUTO MATIC CO., LTD. http://www.dspnauto.com.tw/ep1.htm

[2] J. A. Freeman, and D. M. Skapura, Neural Networks Algorithms, Applications, and Programming Techniques, Add ison-Wesley, July 1992.

[3] S. R. Bhatikar, R. L. Mahajan, “Artificial Neural-Network-Based Diagnosis of CVD Barrel Reactor,” *IEEE Transactions on Semiconductor Manufacturing* 15(1) pp. 71-78, 2002.

[4] C. D. Ruberto, A. Dempster, S. Khan, and B. Jarra, “Automatic thresholding of infected blood images using granulometry and regional extrema”, *Proc. 15th International Conference on Pattern Recognition*, Vol.3, pp. 441-444, 2000.

[5] “LabVIEW Machine Vision and Image Processing Course Manual,” National Instruments, Feb. 2004.

[6] G. L. Ritter, H. B. Woodruff, and S. R. Lowry, “An Algorithm for a Selective Nearest Neighbor Decision Rule,” *IEEE Transactions Information Theory*, vol. 21, no. 6, pp. 665-669, 1975.

ASSET MANAGEMENT CHARACTERIZATION OF THE PORTUGUESE SECONDARY SCHOOL BUILDINGS

Ana C. V. Vieira [a,b] and A. J. Marques Cardoso [a]

[a] *Universidade de Coimbra, FCTUC/IT, Departamento de Engenharia Electrotécnica e de Computadores, Pólo II – Pinhal de Marrocos, P – 3030-290 Coimbra, Portugal.*

[b] *Instituto Politécnico de Tomar, Escola Superior de Tecnologia de Tomar, Departamento de Engenharia Electrotécnica, Estrada da Serra – Quinta do Contador, P – 2300-313 Tomar, Portugal.*

Abstract: During tight financial times, and since the resources dedicated to the maintenance and operation of school buildings infrastructures come mainly from the state budget, the maintenance and operation budgets are frequently among the first cuts. Public schools infrastructures often resent themselves from this philosophy, presenting in some cases precocious degradation, generally as the result of the priority on allocating funds to items that directly affect education. It is important to focus on areas such as maintenance, building systems, safety improvements and technology, leading managers to improve maintenance performance of these organizations, if possible anticipating in time problems and opportunities. In this scenario, maintenance audits provide a framework for organizations to systematically review, analyse and recommend improvements in performance.

This paper presents the results of a study regarding the Portuguese Educational Buildings, specifically the ones concerning Secondary Schools. It identifies some areas needing improvements, as well as the suggestions made to improve maintenance management, regarding management decision making, human resources, information systems and strategies employed in maintenance management, among others.

The paper contains the analysis of the information about the Portuguese Secondary Schools maintenance management that was collected in 2004, with the help of a questionnaire answered by 279 Portuguese Secondary Schools administrations. Along with the information regarding maintenance management, it was also asked school administrations to provide information about the building characteristics and security, both inside and in surrounding areas of the school.

Key Words: Strategic asset management, maintenance audit, maintenance management, educational organizations, maintenance strategies in asset management.

1 INTRODUCTION

According to the Portuguese law for the Educational System, published in 1986, the dimensions of educational buildings must provide the possibility to receive a reasonable number of students, in order to guaranty the necessary conditions for a good pedagogical practice and to promote a true scholar community. Simultaneously, the management of spaces, installations and equipments, human and material resources, as well as financial and administrative management, should contribute for the educational and scholar success of each student [[1], [2]].

Despite the importance of building and infrastructures maintenance and its role in cost control, savings in materials and lifetime enlarging of equipment and facilities, maintenance in service organizations is often still regarded as a disturbing factor only. Since the resources dedicated to the maintenance and operation of school buildings infrastructures come mainly from the state budget, during tight financial times the maintenance and operation budgets are frequently among the first cuts. Public schools infrastructures often resent themselves from this philosophy, presenting in some cases precocious degradation, generally as the result of the priority on allocating funds to items that directly affect education [[1], [3]-[6]].

Bad conditions of equipments may interfere directly not only in the organization economics, but also by reducing the overall availability of buildings while simultaneously interfere with occupants' security. This is why both installations and equipments associated with the operation of school organizations must be maintained in good conditions. In order to accomplish that, fundamental areas such as Fixed Equipments, Building Structure, Safety, Technology and Maintenance must be considered in a well-balanced way [[7]].

2 PORTUGUESE EDUCATIONAL SYSTEM

Portuguese educational system has been changed significantly over the last three decades. The thought of generalising the access to all levels of education appears at the beginning of the seventies (XX century), yet, the growing number of students has not been followed by the necessary investments in installations and equipments nor in teaching staff competency until the end of the following decade. The publication of a new law for the educational system in 1986 [[2]], and the almost simultaneous integration of Portugal in the European Union, helped to change the scenario. The first introduced a new perspective of decentralisation in the educational system and the second one gave the economical support needed to invest in educational building renovation, new constructions and equipments [[3], [8]-[10]].

Paradoxically, and as a result of pressures in order to enlarge the educational offer, the significant growth of new constructions tends to coexist with spare or non-existent budgets available to maintain schools in good state of conservation, resulting in functional expenses restraint [[3], [5], [6]].

The growing amount of funds dedicated to education until the end of the last century conducted to a significant development and improvement of human and material resources. As a result, Portugal presents a public expenditure in education that overtakes the average of the countries members of the Organization for Economic Cooperation and Development (OCDE). However, the national effectiveness does not come to the pattern of OCDE and Portugal must reduce significantly its public expenditure [[3], [8]].

Education is one of the most resource absorbing sectors and thus it is becoming more and more important to improve its efficiency [[8], [11]]. Despite the complexity of the educational sector, it is obvious the need to review management and administration models, applying the principles of the law for the educational system published in 1986, in particular the principle of institutional autonomy [[3], [8], [10]].

It is important for managers to improve maintenance performance of school organizations, focusing on areas such as maintenance, building systems, safety improvements and technology, if possible anticipating in time problems and opportunities [[7]]. To accomplish all of those goals these organizations must be determined to manage their available resources effectively and to seek improvements for increased efficiency, while they must implement a regularly scheduled detailed maintenance plan for all building systems [[5]].

3 INTERNATIONAL OVERVIEW

In countries such as the United States of America, the United Kingdom, Canada and Australia several reports related to state buildings' maintenance have been published [[1], [12]-[16]]. Those reports enhance the relationship between budgeting plans and maintenance activities for both equipments and buildings.

The British Standard Guide to Building Maintenance Management is an example of a standardization document whose structure allows its application both in more complex organizations and in domestic properties. This document emphasises the importance of health and security in buildings maintenance, both by the perspective of the building user and the maintenance management work executants [[17]].

The development of databases with information about frequent problems and maintenance global costs as well as long-term maintenance plans and inspections, are some of the several methodologies used to improve buildings maintenance [[5], [12]]. Among the considered requirements are [[1], [12], [13], [18]]:

- Regular inspections for assessing equipment and structures current state;
- Regular inspections for assessing equipment and structures evolutionary state;
- Long time planning for maintenance operations;
- Preventive Maintenance Plans;
- Periodic Audits of the organizational structure of maintenance management personnel;
- Computerized Maintenance Management Programs or similar;
- Data bases with the most common problems;
- Records of all maintenance activities;
- Data bases with all maintenance costs;
- Energy management.

The development and implementation of suitable maintenance plans, for the full range of building systems, equipments and components, is systematically referred. It is also pointed out that such maintenance plans must be the result of a real and participative commitment of each organisation [[5], [13]-[15]].

Both operation and the implementation of the abovementioned requirements involves funding. One of the mentioned approaches to budgeting for preventive maintenance needs considers 5% of the operating budget to be adequate. Another one sets the budget based on 5% of the buildings present value [[13]].

4 MAINTENANCE POLICY FOR PORTUGUESE SCHOOLS

In the mater of building's maintenance, the legislation published in Portugal is vague and sometimes even not existent, which contributes for the precocious deterioration of some buildings. As an example, Portuguese Laws 38382/51, 555/59 and 177/2001 only state that "*(...) at least one time every eight years constructions should be object of preservation actions (...)*". It is also important to refer that building's maintenance activities in Portugal only represent 4% of the global construction activity, while in the European Union the mean value for this percentage is of 35% [[19]].

Simultaneously, the Portuguese law for the Educational System only states that " (…) construction and maintenance of school buildings and its equipment must stand on an effective regional policy, with a clear definition of competencies for every intervenient, who, in this matter, should have the necessary resources(…)" [[2]].

In fact, the importance of building's maintenance is not yet being considered as a key element of Portuguese school board's mission statements. Several studies concerning the educational sector have been carried out in Portugal, dedicated to Primary, Preparatory and Secondary schools, but all in areas such as pedagogics, politics and administration.

Among the numerous publications of the Portuguese Board of Education, a few consider the operational area of educational organizations and only two were found that focused on the importance of maintenance in educational buildings [[6], [20], [21]]. Those publications were presented as informative documents, suggesting that "*(...) each school board should develop its own operation, maintenance and security manual (...)*" considering each building constructive characteristics as well as the installed equipment and collecting data such as [[20], [21]]:

- Buildings layout;
- Data base with all school assets, including operation and maintenance instructions;
- School facilities operation and maintenance instructions;
- Procedures for school facilities housekeeping;
- Procedures for school facilities and equipment periodic maintenance;
- Specification of technical responsibility for electrical and gas systems;
- Outsourcing contracts for maintenance.

5 PORTUGUESE SECONDARY SCHOOLS

The Portuguese Educational System is a complex structure and so, any suggestions to improve maintenance management, regarding management decision making, human resources, information systems and strategies employed in maintenance management, among others, must be based on the objective identification of those areas needing improvements, as the first step of an overall maintenance improvement process.

In Portuguese Educational System the specificity of the several levels of education refers not only to pedagogical practices but also to spaces, installations and equipments policies management, human and material resources management, as well as financial and administrative management. Therefore, the methodology for studying the Portuguese situation should reflect this reality.

In this context, it was decided to evaluate separately the situation for each level of education, not only in terms of maintenance management, but also regarding building characteristics, building systems, safety improvements and technology. Since it was not available any survey of the Portuguese Educational System, referring to those areas in particular, several questionnaires have been developed, to be answered by each school board, with the aim of collecting information for later analysis.

One of the developed questionnaires, designated by CARMA^{EE} questionnaire, was designed in the framework of the Portuguese Secondary School level of Education, and it was posted to all the 869 Portuguese Schools with Secondary Education, whose administrations where asked to provide information regarding their maintenance management policies, building characteristics and security, both inside and in surrounding areas of the school.

The CARMA^{EE} questionnaire was focused on the data categories presented in Table 1:

Table 1

Main structure of the CARMA^{EE} questionnaire

Data categories	Subjects
School Characterization	School identification and building characterization
	Geographical location
	Functional characteristics and socio-cultural structures
	Transportations and parking facilities
	Human resources
	General information
Security	People safety and assets security
	Fire protection
	Gas infrastructures and equipments
	Safety, health and hygiene
	Electrical infrastructures
Maintenance Management	Human resources
	Documentation
	Annual budgeting and operations and maintenance costs
	Building infrastructures and equipments conditions
	Maintenance strategies
	Large retrofitting interventions

The information about the Portuguese Secondary Schools maintenance management was collected in 2004, with the help of the above mentioned questionnaire, answered by 279 Portuguese Secondary Schools administrations. The analysis of the gathered information allowed the identification of the areas of opportunity in terms of maintenance management enhancement. It is important to note that, for this kind of studies, it is considered to be acceptable a percentage of participation from 5% to 40% and that the number of received answers exceeds 32% of the 869 Portuguese Schools with Secondary Education.

The results presented here are based on the analysis of the answers to the set of questions associated to Maintenance Management and its related subjects, as asset condition appraisals and maintenance planning.

5.1 General overview and budgeting

Portugal is divided in Continental Portugal and two autonomous regions, with specific administrative and fiscal framework: Açores and Madeira. The designation Continental Portugal is used to differentiate the mainland territory of Portugal from the insular territory, which is composed by Açores and Madeira Portuguese archipelagos, located in the Atlantic Ocean. Continental Portugal is divides in 18 districts, considered not only for political and administrative subdivisions but also for public education or health care purposes.

Figure 1 represents mainland and insular Portugal territory, divided in its 18 districts and two autonomous regions.

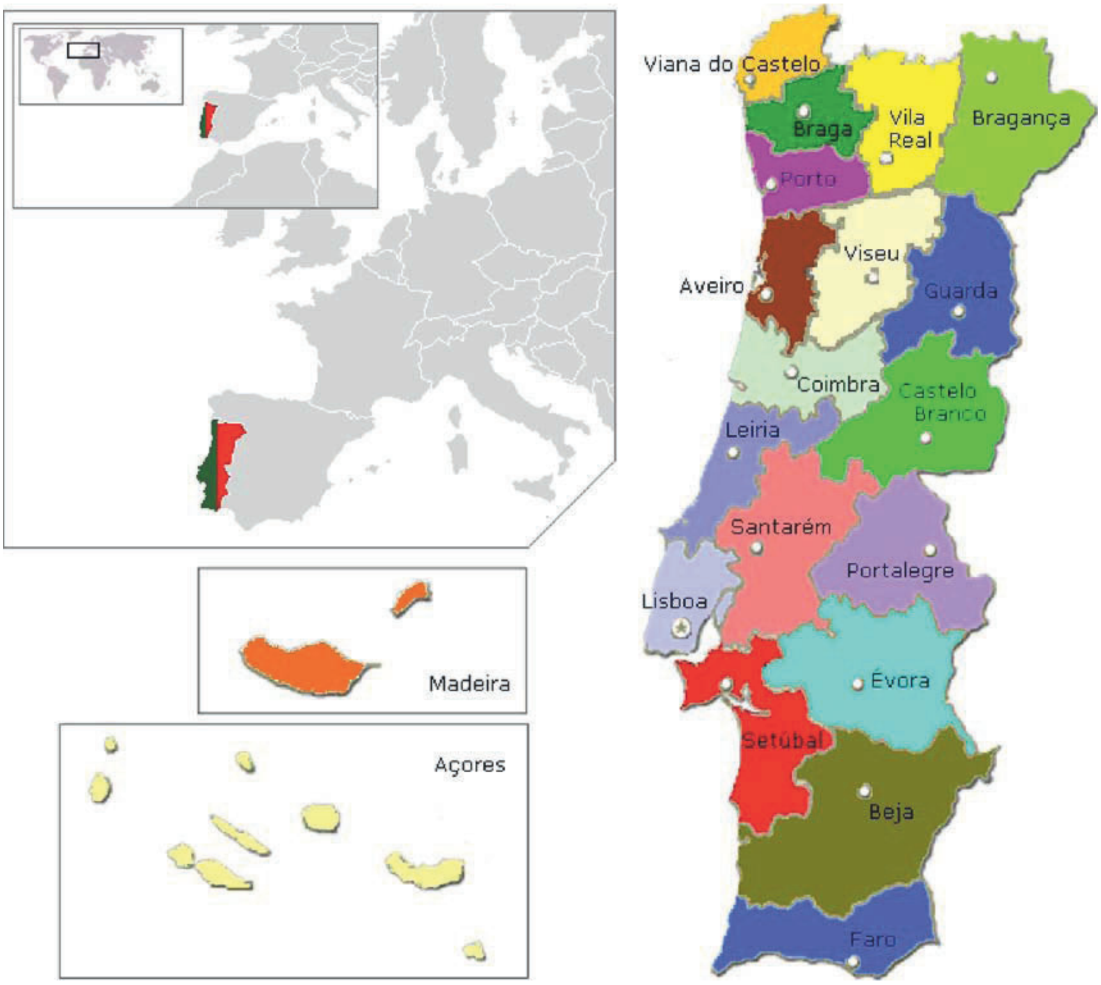
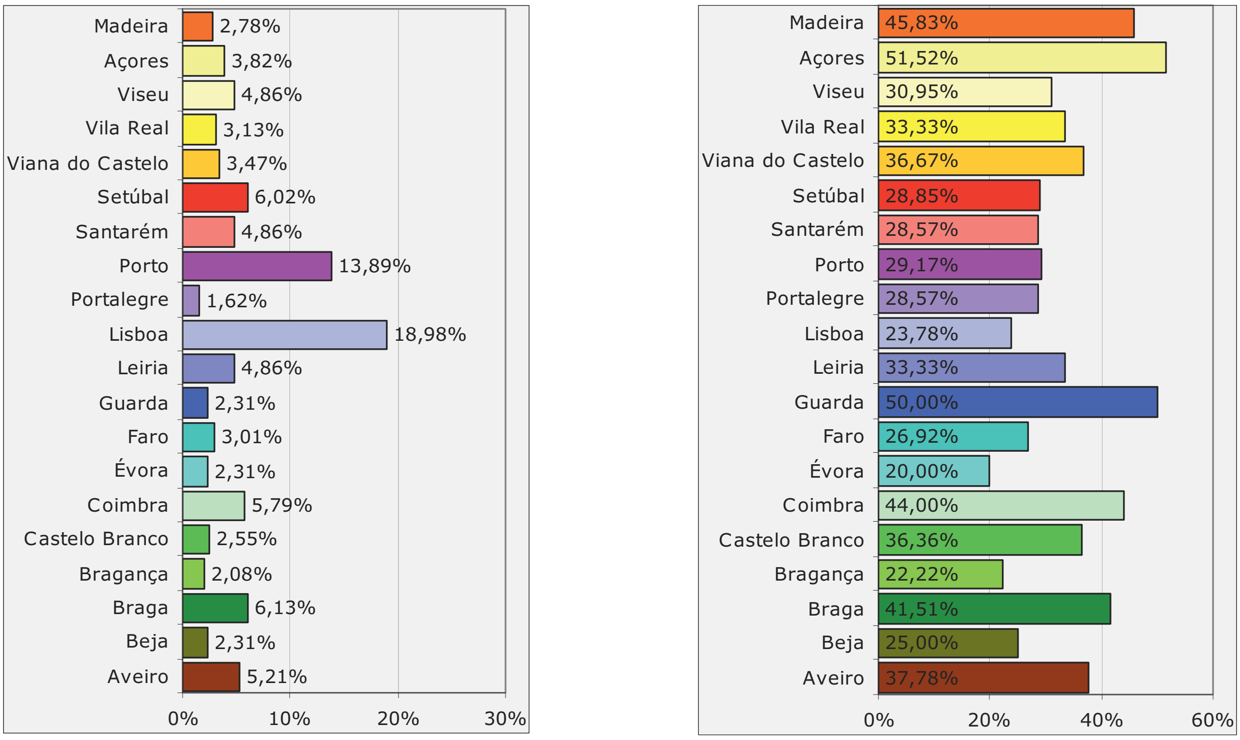


Figure 1 – Portugal in the world, mainland districts and autonomous regions of Açores and Madeira.

As illustrated in Figure 2, among the 20 Portuguese districts and autonomous regions (Açores and Madeira), the highest rate of participations occurred in Açores archipelago with 51,52% responses received. On the other hand the lowest rate of participation (20,00%) refers to Évora district.

Considering the 279 Portuguese secondary schools participants, 47,25% have less than 20 years old and only 12,92% of those schools buildings have more than 50 years old. As illustrated in Figure 3, all the Viana do Castelo district participant schools occupy buildings with less than 25 years old and, on the other hand, all the school buildings in Bragança have more than 20 years old.



(a) **(b)**

Figure 2 – Portuguese school institutions with secondary level of education: (a) distribution of the 869 schools by the 20 districts and autonomous regions under analysis; (b) percentage of participation from the 20 districts and autonomous regions under analysis, considering the 279 answers.

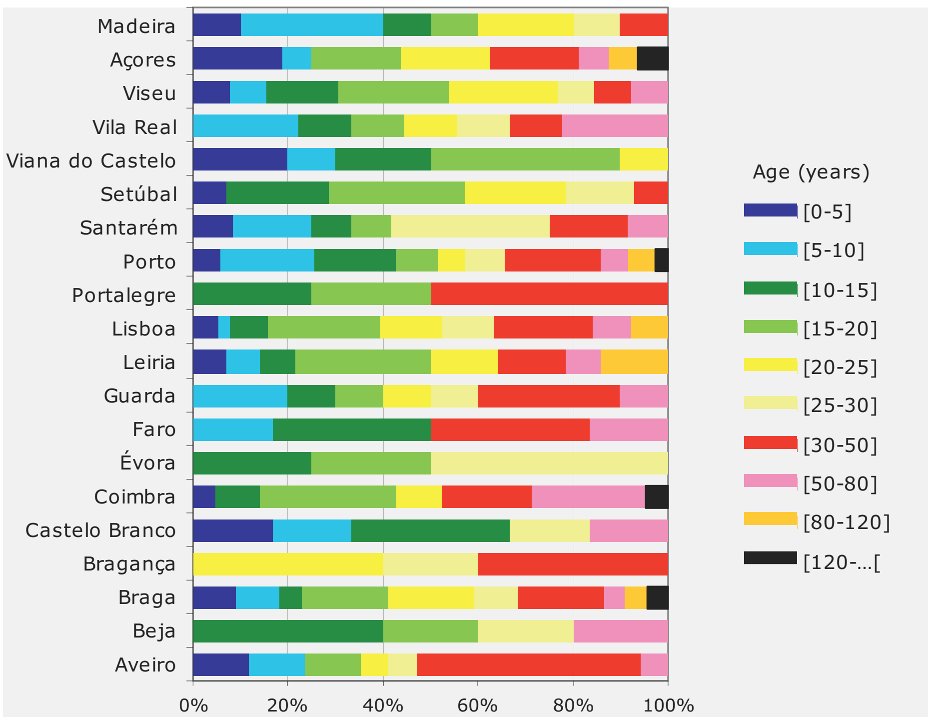
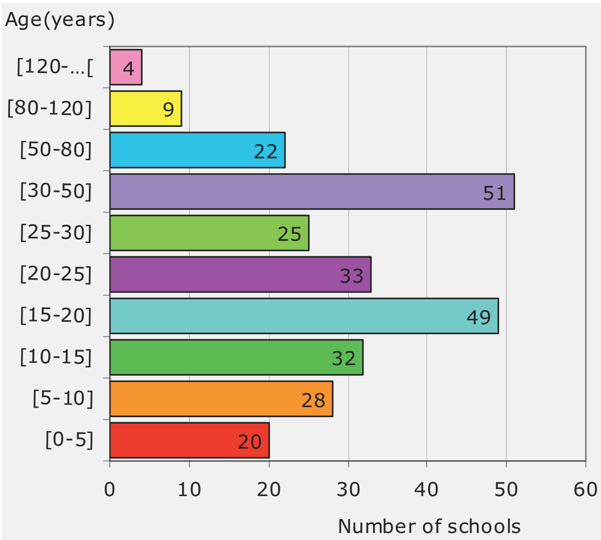


Figure 3 – School buildings distribution by age, considering the districts and autonomous regions of Portugal.

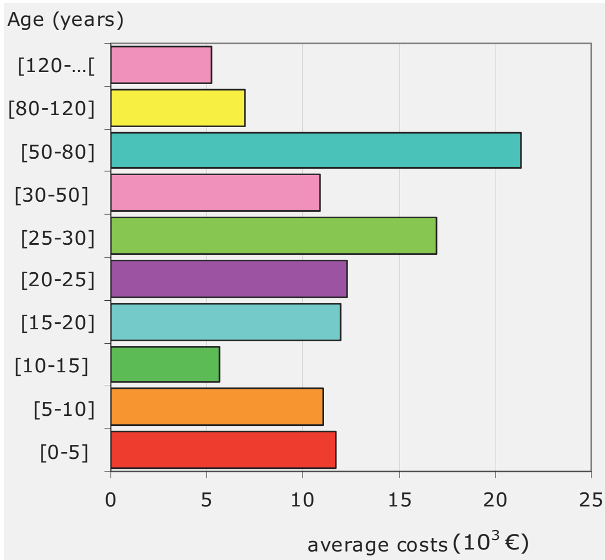
Figure 4(a) illustrates that 39,78% of the school buildings have more than 25 years old and 80 of the 279 participant institutions (28,67%) occupy buildings with less than 15 years old.

The maintenance and operation costs indicated by each school board reveal that those expenses are not only a function of the buildings age and size but also depend on the public or a private administration of the schools. Effectively, in Figure 4(b) the average values for each age interval tend to follow the building age, except for the extremes (below 10 years and above 80 years). The lowest values corresponding to school buildings with more than 80 years and less than 10 years old can be explained by the relation with the buildings areas, as illustrated in Figure 4(c).

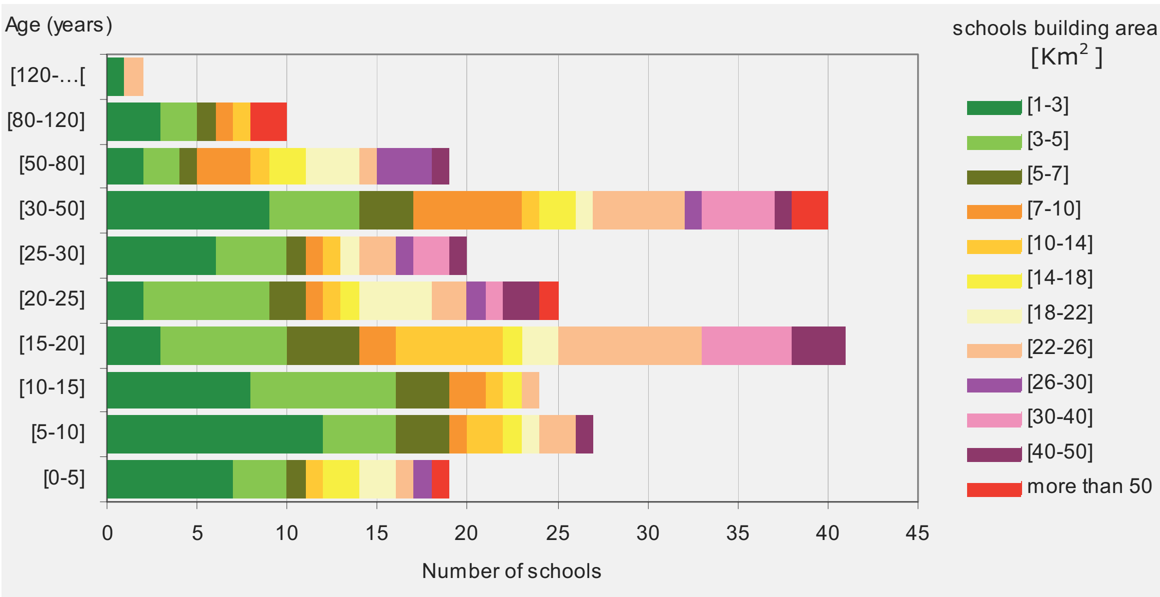
From the 279 participating schools, only 181 refer the amount of money expended in outsourcings, for example in maintenance of elevators and acclimatization systems. Simultaneously, no more than 81 of these schools assumed to have an annual budget for operation and maintenance.



(a)



(b)



(c)

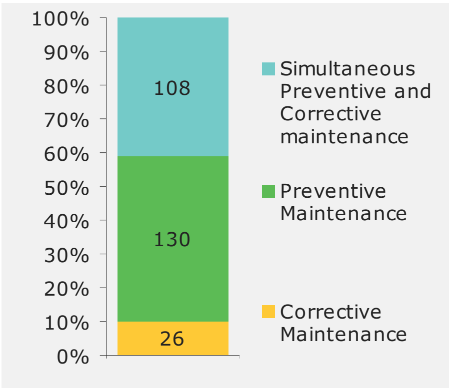
Figure 4 – School buildings distribution by age: (a) number of schools for each given age interval; (b) school's average maintenance and operation costs in each age interval; (c) schools building area in each given age interval.

5.2 Maintenance strategies

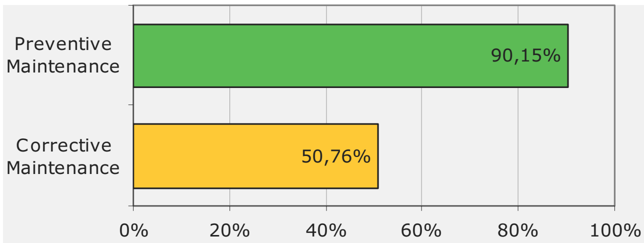
Figure 5(a) presents the number of schools referring to follow only one of the preventive or corrective maintenance strategies and those using both strategies simultaneously. As shown in Figure 5(b) a significant number of those 264 schools (90,15%) refer the application of preventive strategies. Nevertheless, 40 of this 238 schools (16,81%) where not able to identify any equipment or structure maintained in a preventive basis nor the periodicity of the preventive maintenance applied.

For some assets, it was also observed that school boards indicated a time period for preventive maintenance that is not in accordance with the standards. As an example, some schools mentioned a monthly based maintenance of the roofing while others indicated a time period of less than one year (69,56%), despite the 15 years time interval suggested in the literature for the same structure [[6]].

A similar situation was observed with respect to walls. Respectively, 73,33% and 61,9% of the schools indicated time periods of less than one year between preventive maintenance interventions in interior and exterior walls, when the literature usually indicates periods of 5 years between small interventions in interior walls and 10 years between interventions in exterior walls [[6], [12], [19]].



(a)



(b)

Figure 5 – Maintenance strategies: (a) total number of schools referring to follow a maintenance strategy; (b) percentage of the 264 schools using preventive and corrective maintenance strategies.

5.3 Documentation

According to the maintenance policy for Portuguese schools, expressed in section 4, each school should implement its own operation, maintenance and security manual. Several other documents referred in section 3 reinforce the idea that this manual must be developed in order to address the individual and different needs of each school and their systems.

Besides their own manual, documents like manufacturer's literature, legislation or test results may be used by school boards to support their activities. Table 2 shows that 74,55% of the participating schools use support documentation, mainly manufacturer's manuals and legislation (58,65% and 76,92%, respectively).

From the 208 school boards that have answered to this question, only two of them identified the "*Schools Operation, Maintenance and Safety Handbook*", publish by the Portuguese Board of Education, as one of their supporting documents.

The percentage of schools developing their own operation, maintenance and security manuals is very low (22,22%), as shown in Table 2 and every one of the participant schools from the districts of Beja, Bragança, Évora, Portalegre and Santarém and from the autonomous region of Madeira do not have a school own manual.

Effective preventive maintenance programs depend on the feedback from maintenance personnel and the reporting system used, namely with respect to costs associated with preventive maintenance efforts [[13]]. With respect to maintenance management documents, Table 3 shows that 46,59% of the participant schools developed annual documentation, such as budgets (29,03%) and reports for maintenance activities (25,8%). For those schools, 29 (10,39%) guarantee to develop simultaneously Maintenance Plan, Reports and Budget for maintenance, but 23,08% of them refer to follow a corrective maintenance strategy. At the same time, only 59 school boards assured to simultaneously developed documents for recording execution dates, description and costs for every maintenance activity.

Table 2

Support documentation for safety and maintenance activities

Use of supporting documentation	Frequency	Percentage of the 279 participating schools	Percentage of the 208 schools using documentation
Yes	208	74,55%	
School own manual	62	22,22%	29,81%
Manufacturer manual	122	43,73%	58,65%
Legislation	160	57,35%	76,92%
Others	11	3,94%	5,29%
No	51	18,28%	
Do not answered	20	7,17%	

Table 3

Documents for maintenance management developed by the school board

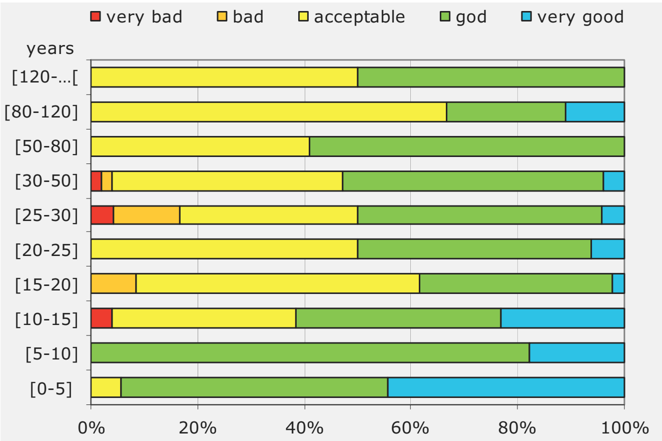
Annual documents	Frequency	Percentage of the 279 participating schools	Percentage of the 208 schools using documentation
Yes	130	46,59%	
Maintenance Plan	77	27,59%	59,23%
Reports	72	25,80%	55,38%
Budget for maintenance	81	29,03%	62,31%
No	138	49,46%	
Do not answered	11	3,94%	

5.4 Building infrastructures and equipments conditions

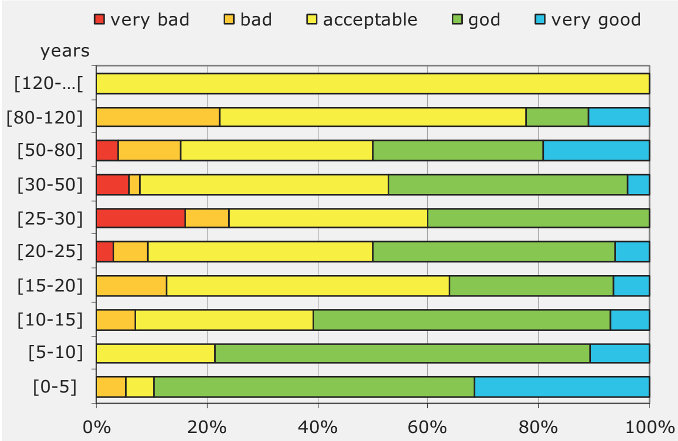
Figure 6 presents two examples of the analysis made for school building's assets general situation.

In the particular case of the interior lighting systems, Figure 6(a) shows that, in general, these assets are considered by school boards administration to be in god conditions. Only three schools have very bad problems with interior lighting but two of them have already asked for an intervention.

The results presented in Figure 6(b) show that almost 20% of those schools with ages between 25 and 30 years old have serious problems with their interior walls. It is important to note that one of these schools considers itself to be using preventive strategies.



(a)



(b)

Figure 6 - Building infrastructures and equipments conditions: (a) interior lighting systems; (b) interior walls.

6 CONCLUSIONS

In order to manage the available resources in an effective way, it is extremely important that educational organizations may have actualized, detailed and accurate information about the buildings, their systems and equipments, and the way they are managed. In this scenario, maintenance audits provide a framework for organizations to systematically review, analyse and recommend improvements in maintenance management performance.

This paper presents the results of a study regarding the Portuguese secondary schools buildings and it emphasises the necessity to develop periodic asset inspections, not only to verify their condition, but also as a support to establish retrofitting maintenance programs.

The study supported by CARMA^{EE} questionnaire lead to the main conclusion that it is very important to review management and administration models, since it is necessary to improve the global effectiveness of the Portuguese educational system, namely by using resources in a most effective way.

Not only school board administrations but also the authorities responsible for educational system management must be aware of the importance of school building's assets maintenance management and so training must be considered. Areas such as maintenance management and organization provides the indispensable technical support for their activities of planning and scheduling maintenance activities, as well as a support for inspection activities, reporting and analysis of the collected data. Additionally, team management, leadership and motivation are a valuable support for school managers to better deal with maintenance personnel.

Unquestionably, school buildings have management special needs but they also present unique development challenges in terms of new studies and applications. Simultaneously, school organizations may have a privileged influence both in individuals and institutions.

When effectively applied, maintenance management and organization allows school buildings to be preserved in good operation conditions and simultaneously guaranties occupant's health, well-being and safety. Therefore, maintenance management and organization will contribute to ensure Portuguese educational system effective operation, reason why it is important to develop not only manuals but also standardization for buildings maintenance in general and for school buildings maintenance in particular.

7 REFERENCES

[1] Vieira, A. C. V., "Maintenance Management and Organization of the Facilities and Installed Equipments of the DEEC/FCTUC" (in Portuguese), MSc. Thesis, University of Coimbra, Portugal, 2003.

[2] Law n.º 46/86 of October 14th, "Educational System Foundation Law"(in Portuguese), Diary of the Republic – I Series, n.º 237, pp. 3068-3081, Portugal, 1986.

[3] Vieira, A. C. V., Cardoso, A. J. M., "Maintenance Audit of a School Building", Proceedings of the 16th International Congress on Condition Monitoring and Diagnostic Engineering Management, Växjö, Sweden, pp. 79-87, 2003.

[4] Vieira, A. C. V., Cardoso, A. J. M., "Some Improvements on the Maintenance Management of a School Building as the Result of a Maintenance Management Audit", Workshop Proceedings of the 17th European Maintenance Congress, Barcelona, Spain, pp. 71-78, 2004.

[5] Vieira, A. C. V., Cardoso, A. J. M., "Maintenance Requirements of a School Building and its Influence on the Maintenance Program", Proceedings of the 17th International Congress on Condition Monitoring and Diagnostic Engineering Management, Cambridge, UK, pp. 533-542, 2004.

[6] Coelho, P. C., "Conservation of Educational Equipments" (Portuguese version of a UNESCO document), UNESCO and Portuguese Board of Education, Portugal, 1988.

[7] Ennis, M., Khawaja, S., "Helping Schools Make the Grade", Maintenance Solutions, February 1999.

[8] Tavares, L. V., "Education – From Growth to Development", Cap. 2 in "To Reform Portugal – 17 Strategies to Change" (in Portuguese), edited by Tavares, L. V., Mateus, A. and Cabral, F. S., Oficina do Livro, Lisbon, 2003.

[9] Costa, H., Nobre, L., Serrano, M., "Portuguese Educational System: Characterization and Proposals for the Future" (in Portuguese), Board of Education, Portugal, 1998.

[10] Delgado, J. M., Martins, E., "Autonomy, Administration and Management of Portuguese Schools – 1974-1999" (in Portuguese), Prospective Evaluation and Planning Department of the Board of Education, Portugal, 2002.

[11] Mateus, A., "Economy – from Structural Adjustment to Growth Rethrowing" Cap. 1 in "To Reform Portugal – 17 Strategies to Change" (in Portuguese), edited by Tavares, L. V., Mateus, A. and Cabral, F. S., Oficina do Livro, Portugal, 2003

[12] Flores, I. S., "Maintenance Strategies –Current Building's Envelope"; (in Portuguese), MSc. Thesis, IST, 2002.

[13] Mearing, T., Crittenden, E., Morgan, M., Guess, G., "Alaska School Facilities Preventive Maintenance Handbook", Department of Education & Early Development, 1999 Edition.

[14] Tasmanian Audit Office/Parliament of Tasmania, "Auditor-General Special Report No. 23 – Managing School Maintenance and Minor Works", November 1997, ISBN 0 7246 46949.

[15] Hauer, J., Bombach, V., Mohr, C., Masse, A., "Preventive Maintenance for Local Government Buildings – a Best Practice Review – report #00-06", Office of the Legislative Auditor – State of Minnesota, April 2000.

[16] Advisory Committee on Building Condition, "Report and Recommendations", Texas Higher Education Coordination Board, April 2000.

[17] BDB/7; "BS 8210:1986 – British Standard Guide to Building Maintenance Management"; British Standards Institution; 1986.

[18] Organization of American States; "Maintenance Manual for School Buidings in the Caribean"; March 1998.

[19] Colen, Inês Flores; "Proactive Maintenance of New Constructions"; (in Portuguese), IST, 2003.

[20] General Board for Educational Administration, Office Board for Educational Equipments Quality, "Schools Maintenance and Utilization Handbook" (in Portuguese), Board of Education, Portugal, 2000.

[21] General Board for Educational Administration, Office Board for Educational Equipments Quality, "Schools Operation, Maintenance and Safety Handbook" (in Portuguese), Board of Education, Portugal, 2003.

DEVELOPING STRATEGIC ASSET MANAGEMENT LEADERS THROUGH POSTGRADUATE EDUCATION

Dr David Thorpe

Faculty of Engineering and Surveying, University of Southern Queensland, Toowoomba, Queensland 4350 Australia.

Abstract: The modern engineering asset manager is required to develop, operate and maintain engineering assets economically and in a socially responsible and sustainable manner. Key issues confronting the asset manager of today include taking a strategic life cycle approach to asset management, meeting user requirements and minimising risks. To achieve this, it is necessary to understand asset life cycle issues including economic analysis and sustainability, be aware of social impacts, understand technological risks and work at the cutting edge of technology. While undergraduate engineering programs can meet this need to some extent, postgraduate education is often the best approach for learning the principles of strategic life cycle asset management. This paper discusses one such example of postgraduate engineering education, the Master of Technology Management, offered by the Faculty of Engineering and Surveying at the University of Southern Queensland. Topics discussed include the rationale for this type of study program, how it addresses the needs of life cycle asset management including risk and innovation management, its approach to learning, and its role in developing strategic technological leaders. The future role of programs of this type in developing leading asset management professionals is also discussed.

Key Words: Asset management, Technology Management, Postgraduate, Education

1 INTRODUCTION

The engineering asset manager of today faces many challenges. The types of assets managed may vary from infrastructure assets like major physical installations and distribution networks, to fixed and mobile plant and equipment such as manufacturing installations and transportation equipment, through to communication networks.

Asset management in such an environment is a complex task that requires application of a combination of engineering, strategic management and financial skills. In the modern world, it is also just as important to achieve social and environmental goals as it is to reach the more traditional technical and economic goals. The modern engineering asset manager is therefore required to develop, operate and maintain engineering assets economically and in a socially responsible and sustainable manner over the life cycle of the asset and throughout decommissioning and disposal.

While undergraduate engineering study programs include engineering management courses, their ability to develop the skills for effective asset management are limited, as undergraduate programs are principally aimed at meeting formal academic training to meet the requirements of professional organisations. Engineering management training in such programs is therefore of necessity at a fairly basic level. Postgraduate programs, aimed at providing sound engineering management training in asset management and related concepts like sustainable development, technological innovation and risk management, can achieve the goal of offering the additional academic development needs by today's integrated strategic asset managers.

It has been in this environment that programs such as the Master of Technology Management (MTM) program at the University of Southern Queensland (USQ) have been developed. As well as taking graduate engineering and technology management education beyond that at undergraduate level, programs of this type aim to develop the skills needed to compete and be successful in the complex world of technology, engineering and entrepreneurship in which the advanced engineering and technology manager will work.

This program was developed out of the recognition that a large number of engineers and other practicing professionals aspire to managerial positions in a technology or engineering environment. It was also recognised that qualified managers of technology play a crucial role in technologically advanced as well as developing societies. Thus it was reasonable to expect that many of these professionals would want to achieve postgraduate qualification in a coursework-based management-focused program [1].

This paper discusses the way in which the programs like the Master of Technology Management can assist to develop strategic asset managers through an integrated program that teaches not only asset management but also explores related topics like technological management and its impact, sustainable development, management of technological risk and other key technology management areas. Initially, the paper discusses the context and issues in strategic asset management. It then explores the options for strategic asset management education, discusses the MTM program and the approaches it uses for teaching, evaluates what it has achieved so far and explores the challenges of the future.

2 CONTEXT OF STRATEGIC ASSET MANAGEMENT

In developing a study package suitable for strategic asset management in an engineering environment, it has been firstly necessary to establish the context for the asset management process. This has consisted of developing an understanding of the types of assets being covered by the program and what the term "asset management" means in the engineering asset management environment; and from this developing the concept of the asset life cycle and the goals that need to be met by the asset manager. This was followed by an evaluation of the challenges faced by asset managers in the delivery of their responsibilities.

While an asset may be broadly defined in a financial sense to mean any item of economic value owned by an individual or corporation, it was decided that in an engineering sense an "asset" would normally mean installations, plant, equipment, knowledge, resources and related items that that support the community, commerce and industry.

Like "asset", the term "asset management" can be also defined in several ways. A common definition of asset management used in highway engineering is "A systematic process of maintaining, upgrading and operating assets, combining engineering principles with sound business practice and economic rationale, and providing tools to facilitate a more organised and flexible approach to making the decisions necessary to achieve the public's expectations" [2]. While not all engineering assets will have the public as asset owner or user, this definition can form the basis of a more general description of many engineering assets by substituting the words "expectations of asset owners and users" for the words "public's expectations."

The life cycle asset management process breaks into the four main streams of identification of the need for the asset; provision of the asset (including its ongoing maintenance and rehabilitation to suit continuing needs); operation of the asset; and disposal of the asset [3]. In performing these tasks, the asset manager is required to not only use sound engineering principles, but also sound business practice and an economic rationale. Modern asset managers also need to take account of social and environmental issues. The strategic asset management leader also needs to look to the future, and therefore as well as applying sound budgeting and engineering management practices is required to consider the long term as well as the short term.

Engineering assets typically serve a number of communities of interest (or stakeholders), and are usually closely liked with the environment of which they are part. For example, physical infrastructure assets are founded on the natural environment and support the economic environment and social environment; and serve the communities of interest of the owner and/or manager, user and the community external to the asset (see Figure 1). A road, for example, is constructed on a natural foundation and is part of the wider environment; supports the economic environment represented by the transportation industry; and serves the road owner, vehicle drivers and other users (such as cyclists and pedestrians), and the community external to the road (such as residents or owners of the properties passed by the road and the wider community whose products they transport). A similar discussion can be developed for other fixed (in location) assets like power stations, water supply and sewerage facilities, and distribution systems. This thinking can be extended to other assets, such as mobile plant and equipment.

The communities of interest to assets each have different requirements of the asset manager, who therefore has to meet a number of goals to meet these requirements. Thus, users expect the asset to have the ability to provide an adequate level of service at the required level of demand; and meet a required level functional serviceability (i.e., condition). Owners and managers require the asset to deliver the optimum service life consistent with the requirements of stakeholders; deliver maximum benefit over the life cycle consistent with other requirements; and operate at minimum life cycle cost, again consistent with other requirements.

Stakeholders external to the asset can be divided into the two sub-groups listed above – the local external community and the wider community. The local external community, which will have relatively close geographical proximity to the asset, will require that the asset satisfies external requirements of interest to that community (for example, environmental and political requirements) to a level acceptable to the community outside the users, owners and managers. Finally, the wider community, which often funds the asset (for example, taxpayers funding a community asset, financiers funding an industrial asset), will require the asset to deliver a whole of life performance (the integral of functional serviceability) acceptable to its stakeholders [4].

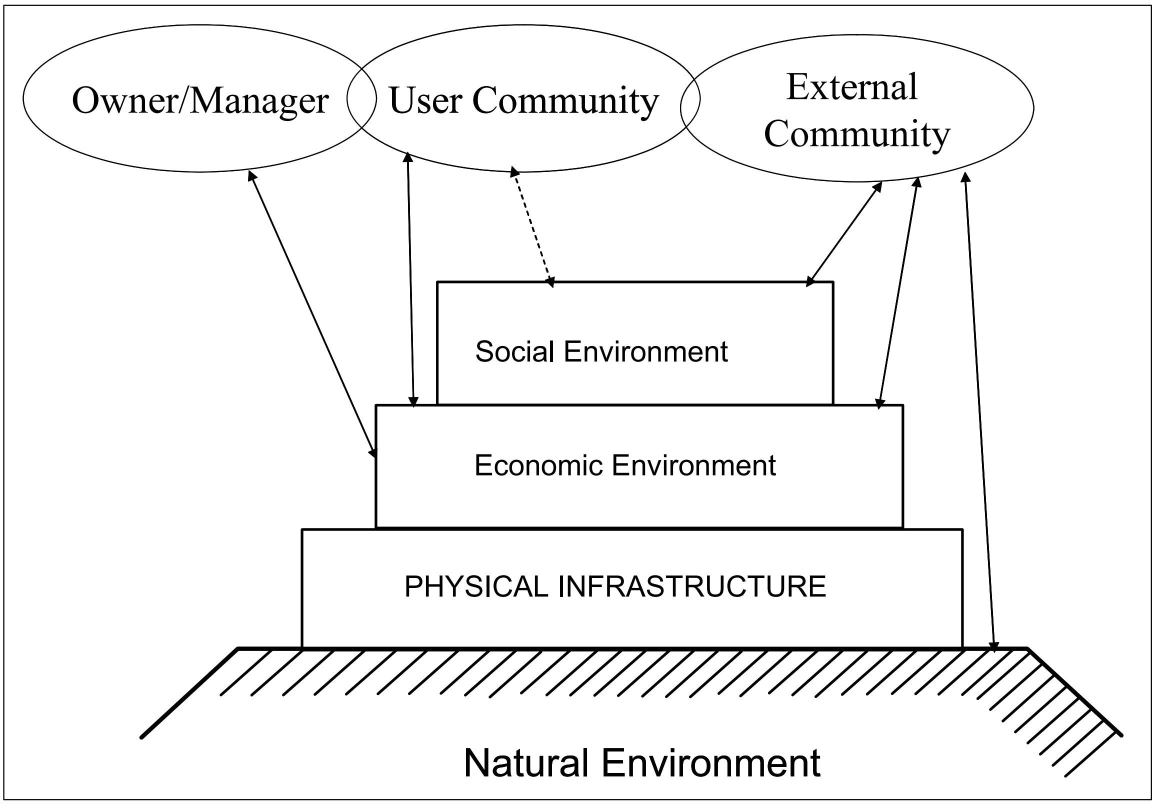


Figure 1: Physical Infrastructure Communities.

Source: Grigg, Neil S. (1988), Infrastructure Engineering and Management, John Wiley and Sons, N.Y., USA.

3 CHALLENGES IN STRATEGIC ASSET MANAGEMENT

While asset managers are required to meet (and optimise) the requirements of key stakeholders, they are also required to meet other requirements. One of these – sustainability - follows from the need to meet social and environmental requirements of the wider community. It is at the core of sound strategic asset management principles that ensure that assets not only meet the requirements of today's society but also the requirements of society into the future. This needs to be considered in practical terms. Johnson [5} sees sustainability is as an ideal state of long-term social, economic and ecological stability, a target towards which we strive, rather than one we expect to reach. In this scenario, sustainable development would consist of striving towards sustainability, while still pursuing production goals and overall economic growth

.

Sustainability has been identified as a desirable quality in engineering leadership competencies [6]. It has also been recognised as an important issue for research organisations, such as the Cooperative Research Centre for Construction Innovation, which has developed a report on sustainable subdivisions [7] and has placed considerable effort into the development of a life cycle assessment tools to evaluate eco-efficiency of buildings from CAD drawings [8]. The social and environmental responsibilities of engineers have also been adopted by Engineers Australia, which in its Code of Ethics has stated that its members 'shall, where relevant, take reasonable steps to inform themselves, their clients and employers, of the social, environmental, economic and other possible consequences which may arise from their actions.' [9]

A common outcome of the discussion on sustainability is that to achieve a sustainable future, any development and management project we undertake needs to carefully consider the needs of future generations, the ability of natural resources to continue to meet the needs of those generations, and what we can do to manage those natural resources to enure that this occurs. This argument can be extended to include all aspects of human endeavour. Thus, when we consider a sustainable future, we should be considering, in additional to environmental sustainability, concepts like social sustainability, economic sustainability, technological sustainability and sustainability in other fields of human endeavour and nature. Examples of this approach include a sustainable and cohesive society, the maintenance of historical artefacts, the ability to economically support the world's population and management of technological knowledge.

Clearly, the asset manager has a significant role in ensuring that assets are planned, designed, developed, operated, maintained, rehabilitated, retired and disposed of in a socially and environmentally responsible manner. Thus, there is a responsibility to investigate and resolve social issues and conduct environmental impact assessments over the whole asset life cycle. An example of this is in the use of some modern asset management materials. Fibre composites, for example, are now becoming used as a strong lightweight construction material, and are also used in applications like bridge asset management to replace timber, which is in quite short supply. While they clearly have benefits from this aspect of sustainability, there is also potential concern with their disposal when the structure of which they are part has come to the end of its useful life. Will such materials be disposed of easily? Should consideration be given to making them biodegradable at the end of their useful life? What other issues will arise? Such considerations require careful thought by the strategic asset manager.

A second key issue for asset managers is risk management. Risk is defined by the Australian/New Zealand Standard for Risk Management as a 'systematic process to understand the nature of and to deduce the level of risk' and risk management as 'the culture, processes and structures that are directed towards realizing potential opportunities whilst managing adverse effects' [10]. Risk so defined can lead to be both positive and negative outcomes.

The process of risk management as defined in the standard consists of seven main components: communicate and consult, establish the context, identify risks, analyse risks, evaluate risks, treat risks, and monitor and review [11]. It is important to recognise the importance of communication and consultation in this process.

An example of the role of risk management in asset management is the abolition by the High Court of Australia of the 'highway rule', which protected road authorities from action being brought against them due to failure to maintain a road, as part of the common law of Australia. This ruling is further explained in the cases of Brodie v Singleton Shire Council and Ghantous v Hawkesbury City Council [12]. It has led impetus to the management of risk in civil engineering asset management and of the development of good asset and maintenance management systems.

In some cases there have been specialised adaptations of risk management principles for a particular requirement. Thus, the risk management process described in the International Infrastructure Management Manual assesses critical failure in terms of a business risk exposure model that combines the costs of the consequences of failure of the asset (for example, repair cost, loss of income, loss of life, injury) and the probability of failure. The approach used in this manual has a strategic focus – 'should do' work that has to be completed within the next five years, 'could do' work that could possibly be deferred for another five years, and 'defer' work that could be deferred for a further five years. It also discusses strategies for risk management, such as risk reduction through capital or maintenance expenditure, emergency response plans, acceptance of some risk and carrying the consequential loss, and insurance against the consequential loss [13].

Other challenges for the strategic asset manager include the importance of seeking and adopting innovative practices in the asset management process, balancing short-term and long-term asset management requirements, quality management; systems management; project management; and management of resources. In particular, the asset manager is required to maximise life cycle economic return on the asset while optimising stakeholder goals, achieve social and environmental requirements, and to effectively plan and manage the asset management process to best achieve these results.

4 DEVELOPING STRATEGIC ASSET MANAGERS

The above discussion has outlined a number of the challenges faced by the strategic asset manager, including the need to optimise the achievement of a number of often conflicting goals, meet the requirements of various communities, address sustainability and other social and environmental issues, manage risk, be innovative, and plan and mange the asset management process for the benefits of key stakeholders.

While experience as an asset manager is relevant in developing these skills, it is important that the effective asset manager be given a solid theoretical grounding in them. From the point of view of the manager of engineering assets, this may be undertaken through specialised study. In particular, the strategic asset manager needs to focus on the larger picture as well as operational issues such as asset condition monitoring and maintenance.

Many engineering undergraduate courses now include courses (or subjects) that address engineering management and the responsibility of the engineer in society, and so partially meet the requirements of the education of the asset manager. This may commence with a compulsory first level course addressing communication issues, followed by courses later in the curriculum that deal with concepts like the engineer and society, and engineering management.

The University of Southern Queensland (USQ), for example, offers the course Principles of Professional Engineering and Surveying, typically taken during the first year of the engineering curriculum, which teaches engineering investigation and report writing. This lays the foundation for further studies in the responsibility of engineers in society and the principles and practice of engineering management.

This is followed by the course Technology and Society, typically taken in the second or third year of the engineer's studies, which builds on this foundation and deals with topics such as the history of technology, sustainability, environmental impact assessment, politics, economics, models of society, social impacts of engineering, a brief introduction to law and some key management issues such as management systems. As this course is offered in the second or third year of the engineering study program, all Bachelor of Engineering Technology (three year) and Bachelor of Engineering (four year) graduates (as well as a significant proportion of those exiting after two years with an Associate Degree) are exposed to key concepts like sustainability, globalisation, economics, and the need for the engineering graduate to relate to society and be innovative in addressing wider societal issues.

A third course, Engineering Management, builds on the principles of Technology and Society to teach management principles and practice for engineers and further develop the engineering student's appreciation of the social environment within which they will practice, and in particular those aspects of the law and ethics pertaining to the engineering profession. An optional fourth level course, Engineering Management Science, which provides the engineer a number of mathematical management tools, completes the undergraduate engineering management curriculum.

These courses, and similar courses at other educational institutions, provide the engineering undergraduate with a foundation in socially and environmentally responsible engineering principles and methods. This can however only be partially achieved in engineering undergraduate curriculum courses, as the large number of courses required to be studied by the undergraduate engineering student and the need to focus in undergraduate study on technical competence mean that there are limitations with developing the skills of effective and efficient asset management during the education of undergraduate engineers. For example, in the University of Southern Queensland engineering curriculum, the courses described above are only about 12.5 per cent of undergraduate study. Their impact is therefore lessened by the other 87.5 per cent of courses required to be studied by the engineering undergraduate.

Further dilution of the education of undergraduate engineers in the strategic management of assets occurs because the types of courses discussed above (with the exception of Engineering Management Science) are primarily verbally based rather than mathematically based, and therefore a number of students (particularly international students whose first language is not English) can find these courses difficult to understand, and hence may not achieve as well in some of them as they would in technically based courses. Students can also perceive that such courses are not really relevant to their future careers as engineering. For example, over the four offers to Semester 1 2005, the Technology and Society course had a drop-out rate of about 15 to 20 per cent, an overall pass rate of about 65 per cent, and only about 10 per cent of students achieving high grades (distinction or high distinction). These figures are worse for international students, to whom English is at best often only a second language.

The other main issue with undergraduate courses are that although they address issues such as social and environmental responsibility, economics, legal issues and management principles and practices, they of necessity cover these areas at very high level and do not really train engineers in the skills required to strategically manage assets. In addition, there tends to not be a very high emphasis on specific training in asset management and maintenance, and in key skill areas like risk management and innovation management.

Undergraduate education therefore achieves a basic understanding of the principles of sustainability and engineering management in graduate engineers, but falls short of providing the real focus on specific skills in particular specialisations such as asset management. Because of this limitation, the best opportunity of providing engineers with the skills of effective and efficient life cycle strategic asset management is through dedicated postgraduate engineering education.

It is therefore contended that postgraduate courses that focus on real tools for practising engineers provide the best approach for developing the required theoretical and practical concepts they require. Such courses must be directly applicable to practising professionals, and must been seen as a real asset to the corporations and institutions employing them. The University of Southern Queensland has consequently developed a number of postgraduate programs that meet these requirements. Several of these have courses suitable for asset managers, while one has a set of related courses that meet the requirements for asset managers outlined in the previous section.

5 POSTGRADUATE ENGINEERING PROGRAMS AT UNIVERSITY OF SOUTHERN QUEENSLAND

Postgraduate engineering education is offered within the University of Southern Queensland by the Faculty of Engineering and Surveying, sometimes in conjunction with other Faculties such as Business and Science. The programs offered include the Master of Engineering Technology, Master of Engineering Practice, Master of Technology Management (MTM), Master of Professional Engineering and Master of Engineering. There are also Engineering Doctorate and Doctor of Philosophy programs available.

In the Master of Engineering Technology, engineers and technologists develop increased technological skills through studying eight courses (equivalent to a year of full-time study) and undertaking a dissertation project equivalent in value to four courses (one semester of full-time study or equivalent), or alternatively studying an engineering management path of equal workload.

Students in the Master of Engineering Practice study a portfolio based program aimed at providing opportunity for experienced technologists to gain recognition equivalent to a four year professional engineering program.

The Master of Technology Management is a quite recent development. It is a coursework program that is aimed at combining technological and managerial skills for practising technologists and engineers. A related program is the related Master of Professional Engineering, which is a partial coursework and partial professional portfolio (and/or research) based program that utilises Master of Technology Management and other courses to develop a higher level of engineering specialisation.

A number of the courses offered in the Mater of Technology Management are also available in the Engineering Doctorate and in other University courses such as the Master of Project Management offered by the Faculty of Business. The management path of the Master of Engineering Technology program consists of a selected number of these courses.

6 OUTLINE OF THE POSTGRADUATE ASSET MANAGEMENT COURSES OFFERED AT USQ

The purpose of the Master of Technology Management is to produce graduates “that are equipped with essential management knowledge and an appreciation of the latest technologies much broader than the initial specialisation”, in order to equip them to manage complex technological or engineering businesses [14].

This program is completed by part-time external delivery mode, over a minimum of six semesters. It consists of twelve courses, four of which are core Master of Business Administration (MBA) courses and eight of which are specialised technology management courses, four of which are considered core to the program. The first specialised technology management courses were undertaken by students in 2004.

The engineering courses in this program are shown in Figure 2.

Core Courses (all to be completed)
Technological Impact and its Management
Towards Sustainable Development
Management of Technological Risk
Asset Management in an Engineering Environment

Other courses (4 to be completed)
Engineering & Surveying Research Methodology
Assessment of Future Specialist Technology
Management of Environmental Technology
Technology Management Practice
Whole of Life Facilities Management
Technological Innovation and Development

Figure 2: Engineering courses in the USQ Master of Technology Management

As previously stated, this program was developed out of the recognition that a large number of engineers and other qualified people aspire to managerial positions in a technology or engineering environment. It was also recognised that qualified managers of technology play a crucial role in technologically advanced as well as developing societies. This latter characteristic, which emphasises the crucial role of the technological manager in society, illustrates the importance of such a program in key technological functions like strategic asset management.

It is primarily aimed at attracting engineers and technologists who wish to develop management skills but want these to be in the field of technology management. They would come from both Australia and overseas, and would be likely to be ambitious and motivated, and see additional qualifications as a means of career enhancement.

In the next section, the key courses related to strategic asset management are described in more depth.

7 KEY ASSET MANAGEMENT RELATED COURSES IN THE MASTER OF TECHNOLOGY MANAGEMENT

The Master of Technology Management program contains a range of integrated courses that address the requirements of the strategic asset manager. Those currently offered are the specialised Asset Management in an Engineering Environment course, Technological Impact and its Management, Management of Technological Risk, Technology Management Practice, Technological Innovation and Development, Future of Specialist Technology and Engineering and Surveying Research Methodology. A course in sustainable development is expected to be developed in 2007.

To show how a program of this type can assist to meet the requirements of strategic management as discussed above, the courses of Asset Management in an Engineering Environment; Technological Impact and its Management; Management of Technological Risk; Technological Innovation and Development; and Technology Management Practice are discussed below in more depth. This is followed by an overall summary of how the program addresses the development of strategic asset management leaders.

7.1 Asset Management in an Engineering Environment

Asset Management in an Engineering Environment is aimed at developing the skills of asset managers in making strategic and operational decisions in the development and maintenance of engineering assets. It recognises that that in the modern world one of the highest expenditure for any government is the cost of developing and maintaining infrastructure.

This course is designed to enhance the ability of managers in making better economical and financial decisions for the construction and maintenance of engineering infrastructure assets. Such decisions, if properly made, will mean better use of resources in obtaining optimum performance and longevity of engineering assets. This course addresses this requirement by taking a strategic view of asset management; discussing the principles of the asset management process and asset management economics; and applying the principles to the asset management process.

Specifically, this course discusses the asset management context as discussed in section 2 of this paper and establishes the need for the asset manager to optimise the requirements of the main stakeholder groups – owner/manager, user (where the user is not the owner), local external community and wider community. Building on this foundation, it then discusses the strategic asset management framework including the asset life cycle and its issues, performance requirements and deterioration. Following a module on asset management economics, the course then applies the principles in asset operations, renewal and maintenance; integrated asset management including investment decisions for both individual assets and asset network; and asset management systems. It concludes with a discussion of future issues in asset management, such as developments in asset management systems, data collection, sustainability and globalisation and management practices.

The course aims to develop a multi-skilled strategic asset manager who is conscious of sustainability and globalisation issues and able to interact with a range of people from many facets of society [15}. A further course, Whole of Life Facilities Management, focusing on the tactical and operational process of asset operation and maintenance, is in the process of development.

7.2 Technological Impact and its Management

Technological Impact and its Management is based on the understanding that that the world of today is one in which there is dynamic change in the creation and development of technology. Therefore, it is necessary for managers of technology to understand the impact of technological development and the ways in which it can affect the society in which we live and the controls necessary to achieve a positive impact on mankind.

This course reviews current technological development, evaluates its impact on the world on we live in, examines the relationship between modern society and technological development, and discusses the role of technological development on wealth creation and business. It also assesses the overall social need to manage such development as well as technology creation, transfer and exploitation. While it does not specifically focus on asset management, it introduces the asset manager to the principles of strategically and innovatively managing technological development in a world that demands increasing social and environmental responsibility in the technological management process.

7.3 Management of Technological Risk

As previously discussed, risk management is a priority area for the strategic asset manager. It is particularly important for the asset management leader who needs to consider a range of risks when managing both single assets and asset networks. This course concentrates on the management of technological risk with respect to both taking advantage of opportunities and minimising the effect of negative impacts. There is a financial incentive, as well as community and corporate responsibility, to achieve the best results from an asset management point of view and to improve reliability. Increasingly, there are statutory requirements to address reliability and safety issues explicitly. In addition, there may be opportunities to manage risks to the benefit of the organisation. Consequently, strategic asset managers need to be aware of the tools and techniques used for the identification, assessment and treatment of technological risks.

In the first part of the Management of Technological Risk course, risk management is discussed in the context of the Australian/New Zealand Standard for Risk Management AS/NZS 4360:2004. In the second part of the course, learners apply risk management principles to technological and engineering projects and processes, and discuss the future of risk management, which is an important tool in evaluating the likely benefits and costs of alternative approaches in developing a sustainable future. Of particular interest to asset managers are sections on risk management in project management; health safety and environmental risks and their management; and a detailed discussion on risk management in asset management, which is based on the principles similar to those of the International Infrastructure Management Manual described in section 3 above.

7.4 Technological Innovation and Development

Technological Innovation and Development builds on Technological Impact and Its Management, and is designed to enable learners to understand the commercial research and development process; appraise the factors which impact on innovation and its development from a managerial point of view; understand and apply the organisational, social and environmental factors which impact on product and process innovation; appreciate and manage the relevant risks; and understand key issues such as intellectual property management and commercialisation. It begins with understanding the concept of innovation and the management of research and development processes, and discusses key issues such as technical risk, intellectual property management, commercial risk and social impact, and the technology transfer process.

From a strategic asset management point of view, this course develops the skills in the asset manager of not only being creative in the asset management process, but also to seek and adopt advances in technology to optimise the life cycle planning, development and ongoing operation and maintenance of the assets for which the manager is responsible.

7.5 Technology Management Practice

Technology Management Practice covers a range of engineering management topics. In particular, it covers the principles of management, project and works management; engineering economics, probability and statistics, law (as related to engineering), contracts and project delivery, accounting and personnel. It complements specialised business management courses taken as part of the Master of Technology Management through teaching specialised engineering aspects of the management process. For example, the project and works management section takes the learner through the basic principles of good project management; then introduces the principles of planning, estimating and scheduling; and discusses procurement, quality management, plant and equipment management, and office management. The contracts module deals with tendering and contract management, and will in future deal with modern forms of project delivery such as public private partnerships and relationship contracting.

This course provides the tools to enable all engineers, including asset managers, effectively manage the engineering function.

7.6 Summary

While this program does not specifically target one particular discipline of the management of technology, it has specific courses designed for the strategic asset manager and a number of complementary courses that use a similar theme of combining technological, economic, social and environmental issues in a strategic framework aimed at developing the technology leader. From the specific asset management viewpoint, it aims to provide courses assist that manager to better implement whole of life strategic asset management, manage in a sustainable manner, manage technological risks, be innovative and apply engineering management principles to the asset management function.

8 DEVELOPMENT OF THE PROGRAM

8.1 Considerations in Development

Programs like the Master of Technology Management are designed to meet the needs of both potential learners and industry. It is expected that most learners will already be practising professionals, located in both Australia and overseas, many of whom will be working in organisations that are dynamically changing and will have different levels of technological maturity. While most students will be engineers and technologists, some (including some asset managers) may have other basic professions. Hence the courses in this program have been written with a view to their understanding by a range of professionals.

Learners in such a study program are likely to have differing needs, and come from a range of organisations. These organisations, and the people in them, will have differing understandings of the concepts in this program, and will also have differing geographic and cultural backgrounds. As an example, engineering asset management is likely to be quite different for a person in a developed country whose main concern is to ensure that an expected standard of service from a particular asset (such as a four lane asphalt surfaced road) is met, and a person in a less developed country, whose main concern might be justifying the construction of a road to a reasonable standard for its expected traffic and community needs. Course material therefore needs to meet both sets of requirements.

Similarly, requirements and expectations may differ across geographical and political regions. Therefore, course development has needed to consider both Australian and overseas practice. For example, Management of Technological Risk is primarily based on the Australian/New Zealand Standard for Risk Management. However, the course developer has had to appreciate that there are also other equally valid risk management methodologies which need to be considered. For example, see Chapman and Ward [16].

Finally, learners are likely to have different levels of access to the courses. In Australia urban centres, for example, most learners will have access to fast Internet access, and therefore they are likely to both expect and be provided with interactive on-line teaching materials. People in remote areas of Australia may well have minimal Internet access, and those in remote parts of the world could have little or no access to the Internet. Assessment (such as assignments) from students in some more remote regions of the world may need to allow for good hand-written assignments as well as the normal typed assignment expected of a person who has ready access to computers, and may need to allow for slower postage from some locations. Therefore, delivery of the course delivery needs to cater for all of these needs and expectations. This is particularly important for asset managers, many of whom will be managing assets in a range of geographical locations, with differing levels of equipment, expertise and access to labour and equipment.

Therefore, when developing the courses in the program, it has been important to be aware of the differing academic and professional backgrounds of learners, their needs and expectations and those of their employers, and the best way in which they can receive course materials and interact with lecturers.

8.2 Challenges in Developing the Courses

Because of the range of student interests and backgrounds undertaking programs like the Master of Technology Management, it has been necessary to develop courses so that they provide sufficient material in the course material to give basic information, while maintaining a strong focus on a balanced management approach. This needs to be balanced by providing sufficient challenge for people who want to learn a topic in depth. There is also a need to meet the requirements of people with differing levels of mathematical background. For example, people who have a bachelor level degree in engineering technology are likely to have less training in formal mathematics than people with a four year engineering degree. Asset managers can come from both groups, and also from people with business and other backgrounds with a lower level of mathematical skill than fully qualified engineers. Therefore, the level of mathematics in the courses in this program been kept to a minimum, subject to ensuring that necessary standards are maintained.

This requirement has been met by guiding learners through the necessary basic process while giving them access to learn the necessary basic quantitative skills (such as the use of statistics). At the same time, the needs of advanced learners have to be considered. Therefore, the courses have been designed to cover the necessary background material at outline level only, and then move fairly quickly to applications. Those learners who need to know the theory in more detail have been directed to appropriate books and websites..

This type of approach has also been used in other aspects of this course, such as the writing of English expression.

Developing courses to meet the needs of diverse learners is best illustrated by the example of developing Asset Management in an Engineering Environment, which is relevant to the needs of the asset manager. Thus, it was necessary to decide the asset management context (as discussed in section 2 of this paper), assess what material should be included in the course, identify the likely profile of potential learners, and decide how course material should be delivered and assessed.

A particular challenge was to decide on the way in which the course should be organised and presented to meet course objectives and to challenge learners. Thus, while the thrust of the course is on the strategic aspects of managing engineering and technological assets, operational aspects of asset management have also been addressed, both to enure that they are fully considered and to lay the platform for a future facilities and maintenance management course that is expected to cover tactical and operational management and maintenance of assets in more detail. In addition, material relating to sustainable development, whole of life focus, risk management, innovation and technological management has been included, at least to an extent that it links to other courses in the program.

A modular design has been employed, both for ease of study and to allow selected modules to be offered as future short courses suitable for possible industry training purposes. Therefore, the first modules that students study deal with theoretical issues, while later modules in the course focus on application of these principles. Current and emerging issues like sustainability, risk management and optimising multiple asset management goals are addressed throughout the course. The course uses guided research – reflection, research into key issues (supplemented by a series of questions to prompt the research process), development of opinions and problem solving. Use is made of on-line research to supplement written material. As with most courses in this program, a textbook is used to supplement study resources prepared by university staff.

To assist learners with mastering the amount of material in the course, it has a number of activities that are classified into differing ranges of importance. The first level is essential tasks, which provide basic knowledge of the course material. Tasks classified as “important” are designed to provide further understanding. They include reading and understanding explanatory course or text material, or undertaking a reading or exercise that aids understanding of the principles being explained. Learners may also optionally further research course material (“background tasks”) or undertake in-depth research to understand it in more depth (“other tasks”). Material in essential and important tasks is examinable and is accordingly structured to enable learners with minimal mathematical background to succeed in the course

9 IMPLEMENTATION OF THE STUDY COURSES

The Master of Technology Management program is in its third year of offer. The first courses, offered in Semester 1 of 2004, were Asset Management in an Engineering Environment and Technological Impact and its Management. These were followed in the second semester of 2004 by Management of Technological Risk. A further three courses were added in 2005. Students enrolling for these courses are from diverse academic backgrounds that include engineering, engineering technology, business administration, technology management and project management.

All course material is provided in written paper form. While there can be some variation, this usually consists of an introductory book, a study book, and a book of readings. The introductory book contains information about the course, the study book contains the course material, and the book of readings a number of readings to aid understanding.

Written learning material has been enhanced with on-line discussion using the WebCT on-line teaching tool. This allows online discussion, notices, and posting of supplementary course material on the web for those with web access (the majority of students). As systems and communications continue to improve, online learning facilities are being enhanced with discussions, quizzes and other interaction with students.

There have been many studies in the literature that support on-line teaching and learning. For example, Macdonald [17] reported on the use of online interactivity in assignment development and feedback in Britain's Open University; Deeks [16] discussed the use of web-based assignments for structural analysis; and Ferris [19] used web-based teaching for management engineering management. However, it is also desirable to be able to meet the requirements of learners in remote locations, and therefore it is expected that there will be a paper based system of study materials for some time.

Because the courses have been offered for a short time only and many have limited enrolment at this early stage, it is too early for statistically valid student feedback. However, comments received have been positive, and the program appears to be meeting student requirements.

10 THE CHALLENGES AHEAD

Educating strategic asset manager leaders requires a combination of understanding technological and economic asset management, the management of sustainability, risk management and innovation management within an overall technological or engineering management framework. Programs like the Master of Technology Management have been developed in an attempt to meet this requirement by endeavouring to combine business and engineering skills into an integrated package aimed at the innovative management of today's and tomorrow's technology assets and organisations in a changing world with changing demands.

Because of its need to be a dynamic program delivered by distance education to motivated professionals, this program has needed a tight focus, suitability for learners with a range of backgrounds and experience, and development within a tight timeframe. As the requirements of stakeholders change, new materials and methods emerge, better systems are developed, and more is understood about asset management, it is important to develop innovative asset management leaders who are strategic thinkers. This requires continuous updating and upgrading of the courses in this program and the program itself.

A significant component of the course adaptation process will be in meeting the needs of world wide approaches and demands in asset management. This will require better understanding and appreciation of international requirements, and of the diverse expectations of many communities across the world, each with differing social, environmental and economic values and needs. Incorporation of this information into the course delivery process is expected to improve the relevance of course material to learners in local communities and at the same time improve understanding of global issues by all learners. Recent feedback from this process is being incorporated in modification to this program and incorporation of some of its courses in other postgraduate engineering programs offered by the University.

One of the key challenges ahead within the whole engineering profession will be the extent to which a traditional engineering undergraduate program, with its strong emphasis on technical training, will be able to produce the graduate who can effectively manage engineering and technological assets. Courses like the Master of Technology Management can help to fulfil this need by offering a diverse postgraduate technology management education that considers current and emerging issues in strategic asset management, with a view to developing leaders in this field who have a combination of business and technological expertise, as well as a solid grounding in current and emerging issues such as sustainable development and risk management..

A future challenge is likely to be the development of industry based programs as well as the individual programs that are primarily currently offered for course work master degrees. In this way, courses like Asset Management in an Engineering Environment are likely to develop to meet a range of specific industry and community based educational and training needs. This will mean that not only should individuals attain the skills necessary for strategic asset and other management, but also that these skills are expected to be transferred to industry and the wider community.

11 CONCLUSION

While there are limitations in the management education of strategic asset management leaders at the undergraduate level because of the need for students at this level to focus strongly on technical issues rather than the broader issues associated with strategic lifecycle asset management, postgraduate education provides the opportunity to develop the broader skills necessary. Through combining business and technological education, and targeting a range of modern asset management requirements such as sustainability, risk management and innovation, postgraduate engineering management programs like the Master of Technology Management have the potential to develop and empower strategic asset managers in the execution of their duties.

Such programs are considered important in the process of educating current and future strategic asset managers in the sustainable management of engineering assets in a diverse, ever changing and increasingly global environment.

While there are a number of challenges to be overcome, education programs of his type aim to provide the opportunity for such managers obtain a level of postgraduate technological management education that enables them to posses the skills required for them to be leaders both now and in the future.

12 REFERENCES

[1] University of Southern Queensland (2002). Submission for the Accreditation of the Postgraduate Certificate in Technology Management and the Postgraduate Diploma in Technology Management and the Master of Technology Management. Toowoomba, Queensland, 11.

[2] Pakkala, P (2002), Innovative Project Delivery Methods for Infrastructure - An International Perspective. Helsinki 2002. Finnish Road Enterprise, 11.

[3] Austroads (2006) What is Asset Management? Austroads Asset Management Newsletter, 6 January 2006, http://www.austroads.com.au/asset/whatisasset.html, viewed 6 January 2006. What is Asset Management?

[4] Thorpe, D. S. (1998), A Process for the Management of Physical Infrastructure, unpublished Ph.D. Thesis, Queensland University of Technology, Australia.

[5] Johnston, S.F. (2003). Sustainability and Globalisation: a positive role for engineering. In A. Brown (ed.), Proceedings of the 14th Annual Conference for Australasian Association for Engineering Education and 9th Australasian Women in Engineering Forum, Melbourne, Victoria: RMIT University, 2003:442-449.

[6] Seek – A Newsletter for Environmental Engineering Educators (2005). Engineers Australia Environmental College, Canberra, Australia, 1 (4), 1.

[7] Miller, A and Ambrose, M (2005). Sustainable Subdivisions:Energy-Efficient Design. Cooperative Research Centre for Construction Innovation, Brisbane, Australia.

[8] Cooperative Research Centre for Construction Innovation (2005). Cooperative Research Centre for Construction Innovation Annual Report 2004-05, Cooperative Research Centre for Construction Innovation, Brisbane, Queensland, p. 54.

[9] Engineers Australia (2000). Code of Ethics, The Institution of Engineers, Australia. Canberra, Australia.

[10] Standards Australia /Standards New Zealand (2004). AS/NZS 4360:2004, Australian / New Zealand Standard for Risk Management, Sydney, Australia and Wellington, New Zealand, 4.

[11] Standards Australia /Standards New Zealand (2004). AS/NZS 4360:2004, Australian / New Zealand Standard for Risk Management, Sydney, Australia and Wellington, New Zealand, 7-10.

[12] Brodie v Singleton Shire Council; Ghantous v Hawkesbury City Council; High Court of Australia, 31 May 2001 [2001] HCA 29; (2001) 180 ALR 145.

[13] NAMS/IPWEA (Association of Local Government of Engineering New Zealand Inc and Institute of Public Works Engineering of Australia) 2006, International Infrastructure Management Manual, Version 3.06, Thames, New Zealand, 3.53-3.76.

[14] University of Southern Queensland (2002). Submission for the Accreditation of the Postgraduate Certificate in Technology Management and the Postgraduate Diploma in Technology Management and the Master of Technology Management. Toowoomba, Queensland, 2002. Unpublished, 7.

[15] Thorpe, DS (2005), Asset Management in an Engineering Environment, University of Southern Queensland, Toowoomba, Queensland, Australia, 7.14.

[16] Chapman, C. and Ward, S. (2003) Project Risk Management: Processes, Techniques and Insights (2nd ed.). Wiley, Chichester, United Kingdom..

[17] Macdonald, J. (2002). Exploiting Online Interactivity to Enhance Assignment Development and Feedback in Distance Education. Open Learning, 16(4), 179-189.

[18] Deeks, A (1999). Web-based Assignments in Structural Analysis. In Proceedings, 11th Australasian Conference on Engineering Education . Adelaide, South Australia: University of South Australia, 1999:246-252.

[19] Ferris, T.L.J. (2003). Something Old, Something New – A Novel Approach to Web Integrated Teaching in Engineering Management. In A. Brown (ed.), Proceedings of the 14th Annual Conference for Australasian Association for Engineering Education and 9th Australasian Women in Engineering Forum, Melbourne, Victoria: RMIT University, 294-301

HIGH VOLTAGE PLANT ASSET MANAGEMENT

B.Finlay [a] D.J.Tusek [b]

[a] *Manager/Advanced Technology Centre, Connell Wagner* finlayb@conwag.com
[b] *Manager/Electrical Connell Wagner, Advanced Technology Centre,.* tusekj@conwag.com
Ring Road, University of Newcastle, Callaghan 2308.

Abstract: There is considerable diversity of practice in High Voltage (HV) plant asset management throughout Australia in terms of geography and industry. This paper reviews the factors that make for good management of high voltage assets with reference to sources such as CIGRE, IEEE, IEC and Australian Standards. Covered are the basics of high voltage plant asset management and the applicability of condition assessment methods. This paper aims to present a clear and rational framework for plant owners to enable them to rate and if necessary to augment their approach to the management of high voltage assets. The content draws on the results of surveys, Standards and the experience of expert practitioners at Connell Wagner's Advanced Technology Centre in Newcastle. The group has been operating for over 48 years in the area of High Voltage plant condition assessment and testing and has been at the forefront of many trends in this field.

Key Words: High Voltage Testing, Insulation Resistance, Polarisation Index, Dielectric Dissipation Factor, Partial Discharge, Dielectric Spectroscopy, Frequency Response Analysis

1 THE AUSTRALIAN CONTEXT

There is a wide diversity of practice in High Voltage Asset Management in terms of geography and industry and those who provide guidance in relation to its management are equally diverse. There are a number of factors at work in the Australian contexts that are having an impact on the way asset management is being carried out. These factors are the disaggregation of the Government owned utilities, the ageing of the plant, the inevitable trend towards higher voltage plants and skills shortages.

In the electricity utility sector with the disaggregation of Government owned centres of expertise the task of asset owners has become more difficult. These centres were an independent source of expertise that provided industry with a reliable and impartial source of information as well as providing well-trained experts able to provide hands on assistance.

The advent of the electricity market has opened up opportunities for investors to come into the market and new smaller and less capital intensive plants are playing a role in the market as never before. These newer plants have been designed with an eye on initial capital cost resulting in very little if any capacity in redundant systems, every item is critical and detailed knowledge of its condition is both vital and a strategic advantage.

As Australian industry seeks greater prosperity the need to increase output is an ever present concern. This pushes plants that were happily working at 415V into the realm where 3.3kV motors and above are now being used to provide more power for the work that has to be performed. The skill set of Asset Managers must be expanded to meet the demands of the plant working at these higher voltage levels and this paper will lay the groundwork for a working understanding.

Whilst many Asset Managers have had experience with switchyards and high voltage switchgear and are aware of their requirements, in general the speed with which experience is being lost is not matching that at which it is being obtained. Thus today more than ever there is a growing need to provide soundly based guidance on aspects of high voltage asset management.

This paper will set out the fundamentals to allow those charged with looking after high voltage plant to achieve the greatest availability of the plant and facilitate sound engineering based asset management throughout the life of the plant.

2 ASSET MANAGEMENT PLANS AND STRATEGIES

Asset Management means different things to different people. Generally speaking Asset Management is the process of ensuring an organisation's assets perform to support the organisation's mission and satisfy the expectations of the various stakeholders in the organisation. Without an effective Asset Management Plan the enterprise cannot take advantage of synergies that exist between different asset classes and cannot guarantee the timely, consistent and effective management of its assets.

There is no one best model for the effective management of the assets of an organisation as the requirements depend very much upon the organisation's culture and external environment. Connell Wagner has found a number of common traits in successful Asset Management systems around the world. These traits include:

- the asset management documentation is well structured and is derived from organisational objectives
- the system is integrated with management systems including quality, occupational health and safety and environmental management systems
- personnel at all levels in the operations and maintenance area of the organisation are involved in its development
- the methodologies are developed through structured techniques such as RCM, or through guidance from technical organisations such as CIGRE's technique for Service Aged Insulation [1] for high voltage equipment
- maintenance systems are reliability based with mature condition monitoring and life assessment technologies
- staff are appropriately trained
- systems include management review of asset management outcomes such as failure, condition and maintenance information
- performance measures are well developed to continually improve the system

The structure of Asset Management systems will vary according to corporate requirements but a commonly used structure includes the higher level organisational Asset Management Plan, a more detailed Strategy for the asset class and individual Work Flow Instructions or Standard Operating Practices for each type of plant.

3 HIGH VOLTAGE PLANT AGEING AND FAILURE

The ageing and failure mechanisms of high voltage assets are many. It is essential for the asset manager to ensure that all possible ageing and failure mechanisms are reviewed including mechanical, thermal, chemical and environmental as well as the usual electrical modes. Typical ageing mechanisms of high voltage assets include:

- Electrical
  - Overstressing of insulating system by external stresses or failure of voltage grading systems
  - Partial discharge due to insulating system overstressing or contamination
  - Partial discharge due to delamination of composite insulating systems
  - Thermal instability of the dielectric in the insulating system
  - Tracking across dielectric surfaces
  - Short circuits between windings and winding components
- Mechanical
  - Mechanical failure due to forces from short circuit currents and vibration
  - Mechanical failure of insulating systems due to ageing
  - Failure of mechanical components due to cycling, creep and fatigue
  - Failure due to vibration or resonance
  - Failure of rotating component bearings
  - Failure of lubricating systems
- Thermal
  - Overheating of contacts/ connections/ joints
  - Thermal ageing of components
  - Failure of cooling systems
  - Ineffective maintenance of cooling system efficiencies
  - Ineffective maintenance of cooling medium condition (eg insulating oil in transformers)
- Environmental / Chemical
  - Oxidation
  - Hydrolysis
  - Corrosion
  - Contamination by ageing products including acids and ozone

From these failure modes, the maintenance and testing methods can be developed using a structured approach such as that outline in CIGRE's document on Service Aged Insulation [1]. This process describes the method used to develop preventive strategies from an analysis of dominant failure modes, ageing processes, influencing factors and available diagnostic techniques. One should never forget that improper maintenance would be a multiplier on any natural process of degradation.

4 LIFE CYCLE

The management of assets is improved if the asset management system takes into account all of the stages and cycles that the plant will experience through its life. These well defined stages each with their own specific risk benefit from particular attention and represent a junction at which it is judicious to qualify the assets' condition. If this is not performed at these stages then a valuable and relatively inexpensive opportunity is lost and no further insight into the risk is obtained. If these stages coincide with contractual boundaries then there is a transfer of risk from one party to another.

These stages are listed below,

- Specification
- Design
- Construction
- Factory Testing
- Delivery
- Installation
- Commissioning
- Acceptance
- Operation
- Maintenance
- Refurbishment/Life Extension

At any one of these stages the decisions that are made or events that occur may in the end have a profound impact on the life expectancy of the plant. Plant that would otherwise readily last for 30 years or longer may be reduced to scrap in a matter of weeks or months due to seemingly small and inconsequential acts or omissions.

Apart from the first item, the next seven items can be covered off by appropriate contractual mechanisms and ensuring that the service provider is appropriately qualified and experienced to perform the work. A most important element in plant life management is the effective management of the vendor's warranty. It is vital that appropriate plant tests are included in the contract with reference to Australian and International Standards appropriate to the plant. Allowing the contractor to specify its own testing standard and pass/fail criteria is negligent.

There must be an enforceable mechanism that assures that the manufactured, delivered, installed and commissioned plant not only meets the immediate requirements but will last well beyond the warranty period and be free from defects that will cause premature failure. Contracts commonly require that products are supplied free from defects, yet do not then specify appropriate tests, Standards and pass/fail criteria that can demonstrate that this is the case. A test prior to warranty expiration is the last opportunity to ensure that the buyer is not burdened with the supplier's risk.

After being placed into service the asset needs to be monitored with a regime appropriate to the asset's class, production impact and safety issues. If not before, then certainly at major outages, tests should be arranged to determine the state of the plant, gauge the requirement for maintenance and determine the stage of degradation or ageing.

It can be stated that not withstanding random failure or overload, that the vast majority of plant that has,

- been properly acceptance tested, as well as tested before expiration of warranty,
- has had applied a monitoring and testing program appropriate to the asset and its environment, and
- is operated within its rating,

should not fail from electrical causes without sufficient warning to enable the safe and timely removal of the plant from service. In fact such a regime when applied to HV motors with epoxy mica insulation would often give many months or years of notice for the vast majority of failure modes.

One should always consider plant that is not immediately required for service and is placed into storage as is often the case with rewound motors in a production constrained plant. The asset may require some degree of monitoring whilst in storage but if it is placed into storage without acceptance testing the owner is taking on all of the risks associated with the plant and absolving the vendor of any practical form of liability. The cost of doing this can be considerable if the plant value or down time cost is high.

It is worth remembering that all products come with defects, whether the defect impacts on plant life is an open question. The only tools available to identify the risk of defects that will greatly reduce plant life are initially contractual and then through condition monitoring throughout the life of the plant.

5 HIGH VOLTAGE CONDITION ASSESSMENT TECHNIQUES

Test and assessment methods are varied in the nature of equipment, required user expertise, ease with which results can be interpreted and diagnostic value. Below is an overview of some of the more common tests associated with high voltage plant condition assessment and testing. It is important that any person who is to effectively maintain a high voltage asset gains some understanding of the nature and effects of partial discharge of which an overview is included below.

It is noteworthy that high voltage plant is often tested above its normal operating level in accordance with the relevant Standards and guidelines and that this is a normal condition even for routine tests. The purpose of HV testing methods is to expose signs of degradation before they become active under operating conditions where they can lead to catastrophic failure. The desire to test at voltages significantly less, whilst generally well intentioned is not motivated by the science, but rather concerns to do with the potential for failure of the plant under test, lack of appropriate test equipment or the economics in bringing to bear the equipment required to do the test. It can be stated that people experienced in HV testing go about their business in a cautious manner, looking for signs of impending failure to prevent plant damage.

5.1 Insulation Resistance (IR)/ Polarisation Index (PI)

Measure the DC resistance of insulation, typically to earth. Used to determine the basic state of insulation with the higher the value the better, very temperature and humidity dependant. Will show up contamination and moisture ingress and low resistance paths. Polarisation Index is the ratio of the 10 minute to the 1 minute IR. It is an additional measure of moisture and contamination, values above about 2.0 are generally acceptable. Depending on the nature of the plant low values can have different implications. The test is at DC thus electric stress is not distributed throughout the insulation with the profile it would have under AC conditions. Depending on the voltage levels to be used, this may impact on interpreting or determine how and to what level the test is performed.

5.2 DC Ramp Test

A DC voltage is applied to the insulation relative to earth and is raised very slowly whilst noting the apparent resistance, a change shows up flaws in the insulation. This test is generally used to determine ground wall insulation properties of windings in motors and generators. Tends not to be used if a fully rated AC testing capability is available.

5.3 Winding Resistance / Circuit Resistance

Basic measure of the continuity and bulk DC resistance of the winding. Requires temperature correction and 4 wire measurements. Provides phase balance check (<1%).

5.4 Winding Ratio

It is a requirement of the Standards that the measured ratios are within 0.5% of the values indicated on the nameplate. This is a requirement mainly for operational reasons and not related to condition assessment although a change in ratio from factory values would most likely signal a fault. The test will find errors in connections, particularly after On-Load Tap Changer maintenance, which is a high risk time.

5.5 Dielectric Dissipation Factor (DDF)

Measures the AC loss when voltage across the insulation is raised. The characteristic of the relation between DDF and voltage is a good indicator of the condition of the bulk insulation. It will detect voids and corona, which can point to problems in areas such as,

- Electric stress levels and clearances
- Contamination of windings and bushings
- Delamination of windings
- Burning of insulation or other such damage
- Thermal Runaway
- Ageing of the bulk insulation system

In general the higher the value of DDF for the same voltage level, material and temperature, the poorer the condition of the insulation. The measurement is temperature dependant and very susceptible to surface contamination and humidity. The test is also very good at indicating thermal ageing and thermal runaway in insulation. Thermal runaway is a condition where insulation locally develops a negative temperature coefficient and the increased heat that it develops then further reduces the resistance, which in turn increases current flow in the insulation and so on until the insulation is destroyed through overheating initially and then possibly a fault. This condition generally affects certain paper-oil insulated systems such as those found in transformer bushings and instrumentation transformers.

In good insulation the voltage level at which the slope of the curve of DDF vs voltage changes significantly upward (the “knee” voltage) should be at as high above normal operating voltages as possible. As the plant ages this knee voltage reduces until there is significant DDF even at operating voltages. At this time the asset is often considered to be at the end of its economic life.

5.6 Capacitance

The capacitance of an object is a fundamental quantity and variations in capacitance with voltage are small in good insulation. This is generally measured at the same time as the DDF. A departure from the known voltage coefficient or a change in the absolute value of the capacitance can be indicative of insulation damage or shorted elements.

5.7 Partial Discharge Testing (PD)

Partial discharge (PD) testing is by far the most insightful method of determining the condition of a degrading insulation system. The measurement works by detecting the small electric current pulses that occur as part of the insulation degradation process. A partial discharge is an electrical discharge that does not bridge the insulation system. When the PD occurs in air it is termed corona. These discharges are transient breakdowns that allow charge to migrate from one location to another in a manner that does not occur at lower voltage levels. All practical materials will undergo partial discharge at some voltage and then as the voltage is increased sufficiently the electric stress will be high enough to flashover the insulation and cause a solid fault.

PD measurements can reveal the,number of insulation degrading voids present

- size of individual voids in energy terms
- distribution of voids within the insulation
- voltage and conditions at which the voids become active
- whether the discharges are to earth, air, oil, floating objects or within the insulation material
- state of degradation of the insulation system
- bar bounce
- clearance issues

The absence of PD at test voltages is the strongest indication that the insulation system is in good condition and is not suffering degradation from the eroding action of partial discharges (at the time of measurement). These measurements are made on all types of insulation systems including insulating materials, power and instrumentation transformers, motors, motor coils and assemblies of coils. Healthy insulation generally exhibits no or low PD at or above its normal operating voltage in accordance with its insulation type. In a PD test the source voltage is raised to a point where PD activity starts (inception) and is then lowered until it ceases (extinction). Similarly to DDF, when the insulation is in good condition both inception and extinction levels are well above operating voltages. In such circumstances if a transient such as a switching or lightning surge initiates a PD it will quickly extinguish of its own accord and prevent further degradation. As the insulation ages and if a fault develops then inception and extinction voltages track downward until even under normal operating conditions PD's are actively degrading the insulation.

PD tests are the most stochastic of the condition monitoring tests, varying widely both in magnitude and temporally. The same test repeated under the same conditions but separated in time may produce very different results. PD's can require a great deal of experience to interpret even thought the basic principles are easy to understand.

To gain the most benefit from PD testing the test voltage needs to be in excess of the normal operating levels with reference to phase-phase and phase-ground voltages or there is little benefit from the test. If the test is done at operating levels or below it only indicates that the plant is discharging at normal operating voltages which generally means it is near or at the end of its economic life.

The presence of PD is sometimes determined by non-HV test methods such as, radio interference, ultrasonic detection, ozone or other gas measurements. The presence of PD in air will generally result in corrosive ozone and nitric acid, which will corrode structural and insulating elements of the plant, this can lead to either mechanical or electrical failure of those elements. Borescopes can be used to look into internal insulating structures of motors when PD has been identified by test and it is considered prudent to visually inspect the insulation.

5.8 Ultrasonic or Acoustic Emission Detection (UES or AES)

This method can be used to indicate the location of discharges in air or oil. When combined with location methods on transformer tanks, it is commonly used in conjunction with PD measurement equipment to localise the position of a source of strong discharges. On the basis of this location and information from internal drawings a decision about the severity and urgency of the problem may be able to be made. Not recommended for shell type transformers and is less successful for transformers with extensive flux shunts.

5.9 Applied and Induced Tests

These tests are the most basic HV proving test and demonstrate the insulation's ability to withstand the rigours of the service environment from a voltage standpoint. The test involves applying a voltage at a level and for a time prescribed by the Standard, in the indicated configuration. The test is a go/no go test. It can be carried out with a PD test at the same time to add additional diagnostic capability. For transformers this is a routine test.

5.10 Ring Flux Test / Electric Core Imperfection Test (EL CID)

The Ring Flux test is used to test for damage to the laminations of stators. It requires about 80% of operating flux to be induced into the stator and a thermal imaging camera or some other temperature measurement device is used to highlight hot spots in the core laminations. Temperature is a major ageing and risk factor for the winding thus ensuring that there is no excessive temperature rise is important.

The EL CID test is a small signal variation of the Ring Flux test. A special instrument is used that measures the loss associated with a flux of approximately 3%-4% of the rated level. In this case there is no temperature rise of significance to measure. A loss component is measured using a small search coil in contact with the laminations. The measured values are related to predicted temperature rise for the fully fluxed condition.

5.11 Recovery Voltage Measurement (RVM) / Dielectric Spectroscopy (IDA) /Polarisation Depolarisation Analysis (PDA)

These methods which can nominally be regarded as equivalent, are generally used to determine the equivalent paper moisture of transformers and cables [2]. This is an important parameter as moisture in the insulation accelerates ageing. The three methods are different approaches to determining the same parameter. For most practical purposes they can be considered as equivalent although certain approaches have advantages under particular conditions.

5.12 Frequency Response Analysis (FRA)

Generally applied to transformers, FRA is a method of determining the impedance or transfer function between different elements of the windings. This captures the electric and therefore the geometric relationship between physical elements of the windings, core and earth. Any deviation from previous results, not attributable to magnetisation of the core, is indicative of a change in those relationships and possibly a shift in the winding. A loss of winding compression or alignment is generally a precursor to in-service failure of the transformer [3].

5.13 Dissolved Gas Analysis (DGA)

If you own an oil filled power transformer or instrument transformer then by far the best value for money is DGA and the science is mature although with the advent of vegetable based insulating oils new science needs to be developed. The oil in a transformer is a time capsule that not only holds information about the past but also enables prediction of the future. DGA is used to find evidence of all manner of maladies such as,

- arcing
- overheating
- **moisture**
- ageing
- paper or metal involvement in degradation
- need for oil reclamation

The virtue of DGA is its ease in sample gathering compared with the high tech requirements of the laboratory that needs to process the sample. It is easy for plant owners to facilitate their own sampling program and then simply send the sample to the lab by courier.

In looking at DGA one must mention oil moisture content. Moisture in the oil is implicated in many accelerated ageing and degradation mechanisms, it cannot be stressed enough that if you are not looking after oil moisture content you are throwing money away. There is no hard and fast rule but it is generally accepted that 20 ppm is starting to get high, 30ppm requires action.

5.14 Furan Analysis / Degree of Polymerisation (DP)

Furans are compounds released from paper into the transformer oil when the paper is degraded during operation. Given that most power transformers contain a substantial amount of paper in their HV elements then the condition of the paper is an important concern. The Furans are measured because they are a non-invasive way of estimating the Degree of Polymerisation (DP) of the paper. The DP is an average measure of cellulose chain length that makes up the paper. This average length is an important parameter in estimating the mechanical properties of the paper. Paper with a DP of less than about 200 has no mechanical strength and will crumble away. Thus as the paper ages the equivalent DP will track from values close to 1000 downward. The rate at which this occurs depends on the operating conditions and the success in keeping up preventative maintenance practices such as keeping oil moisture low. The relation between Furans and DP as well as the notion of bringing a complex variation in paper properties throughout the transformer down to one number is an approximation, the key to using the information is the trend in values over time [2].

5.15 Impulse Tests

Impulses to simulate lighting and switching surges are applied to the plant to determine if the insulation system can withstand such conditions. Due to the nature of the test equipment, impulse tests are generally carried out in a laboratory or at the OEM's test facility.

5.16 Thermography

This is not a test as such but a good component of any HV plant asset management strategy. Used primarily to discover hot spots, it highlights poor connections or plant that is thermally abnormal which can alert to control problems or accelerated ageing problems.

5.17 Visual Inspection

This low tech and hands on approach to asset management is an important part of any strategy. Many forms of degradation or problems that are sure to lead to degradation or failure can be identified through simple visual inspection. Locating of leaks, low oil levels, damaged components, water ingress, noisy or abnormally sounding components are just some of the potential finds.

The basics are summarised in relation to their plant applicability in the table below.

	IR	PI	WR /	DC	DDF	Cap	Ratio	PD	UES	Applied	Induced	Ring / Flux	RVM / IDA / PDA	FRA	DGA	Furan	RSO	Impedan
Motors												X						
Generators												X						
Power Transformers													X					
Instrument Transformers													X					
Cables													X					
Switchgear																		

6 CONTINUOUS ON-LINE MONITORING SYSTEMS

Continuous On-Line Monitoring Systems (COLMS) have been around in one form or another since very early times, e.g. over-current circuit breakers and protection relays. In these forms they have proven to be very reliable and effective. COLMS as applied to monitoring failure mechanisms in the insulation of high voltage electrical systems are far less reliable and effective then their protection counterparts. This stems from the fact that the protection systems work on mathematically well understood and readily verifiable principles. The insulation monitoring systems are not amenable to the same robust detection methods and are also often very adversely affected by interference. In insulation monitoring applications often a breakout from a trend is as important as the absolute value.

COLMS are a great benefit in certain circumstances in particular when there is a need to use the full economic life of a plant when a known degradation or failure mode is active. They almost invariably suffer from a lower level of reliability than the plant that they are monitoring and thus can be an expensive alternative to time based monitoring systems in many cases in particular during the early stages of a plant's life. Monitors that are improperly applied or operated can at best result in considerable expense in unnecessary additional testing of the plant, at worst they can fail to perform their basic task of monitoring and lead to a false sense of security putting at risk the plant and personnel.

Given the expense of COLMS and the relatively poor reliability the choice of an on-line system for high-voltage plant should be made when plant is critical or has a known defect likely to lead to catastrophic failure. If appropriate steps were taken from contract to acceptance testing and then condition monitoring there should be ample indication of incipient faults to better support the decision of where to install on-line systems to nurse through a deteriorating plant. It may be cost effective to consider the installation of a provision for a COLMS through the installation of measurement transducers and connection points rather than a complete system.

It is important to understand that on-line systems are not a substitute for HV testing of plant. The failure modes of high voltage insulation systems are subject to considerable non-linear behaviour and early warning of many types of failure modes can only be obtained through high voltage testing at the appropriate test levels as specified in Standards and just to the convenience of the person performing the testing. Acceptance and warranty testing should only be performed though high voltage testing methods. On-line systems can only show the presence of a condition when it is already active at the operating condition of the plant and has limited early warning capability. In motors and generators with epoxy mica insulation systems, there is generally a considerable warning period of months or years of insulation degradation and on-line systems are best applied to the final stages of the process to ensure that the full economic life is obtained and capital is conserved.

Mechanical failures are in a different category and many types can lead to failure or significant loss of life in minutes or hours, eg, bearing or lubrication failure. The proven methods of vibration and temperature are invaluable in these instances. Of course there is always the need to have trained staff that can act rationally on sensor information to preserve the plant and ensure safety.

7 STANDARDS

The paper has discussed the need for reliance on Standards, the value of which cannot be understated. Standards come about through a process that seeks input from OEM's, industry practitioners and academic circles to reach a majority agreed position on those matters that the Standard seeks to address. As it is the result of a committee process the Standard generally represents an average opinion and depending on the way that the Standards Committee is formulated may result in a bias in the Standard to the views of the dominant group. OEM's have always found it in their interests to be well represented on Standard's Committee's to ensure that impractical requirements do not become part of the Standard. Thus a Standard can be considered to be the distilled wisdom of various interested and authoritative parties on what levels of performance and test results should be able to be met and represent good product.

With this view in mind, the Standard should be the minimum acceptable level. An offer of less than this is an offer to transfer risk from the seller to the buyer. Whilst many failures to meet Standards can be accommodated by extended warranties care should be taken to account for the total cost of plant failure, reduced life and increased monitoring costs. This can only occur if there is an understanding of the nature of high voltage insulation systems, their failure mechanisms and monitoring.

Failure to specify conformance to Standards for products at the contractual stage is a major risk management failure for a business. The only context in which this should occur is when there has been a rational assessment of the additional costs of compliance with Standards, taking into account availability of alternatives, expected cost associated with failure and failure frequency. Otherwise there is no accounting for risk. How many times has a venture of considerable economic magnitude been brought to a halt due to the failure of an element sourced at the lowest quote?

With reference to High Voltage plant items the following is a partial list of Standards that can be referenced,

Rotating Electrical Machines	AS1359, IEC60034, IEC34, IEEE 432, IEEE 56
HV Test Techniques	AS1931, IEC60060
Partial Discharge Measurement	AS60270, IEC60270
Power Transformers	AS2374
Switchgear	AS2560
Bushings	AS1265
Voltage Transformer	AS1243
Current Transformer	AS1675
Cables	AS1429
Shunt Capacitors	AS2897

These Standards often detail factory and acceptance tests that enable vendors to demonstrate the quality of their products as well as often indicating how to condition monitor many aspects of the plant. Plant that can meet the requirements set out in Standards will start service in a more robust condition than plant that does not do so and will generally have a greatly reduced risk of premature failure.

8 SUMMARY

High Voltage Plant assets have their own requirements for effective management that are above and apart from those of their low voltage counterparts. Any enterprise that follows the process of establishing an Asset Management Plan (AMP) which details and addresses these needs within the context of the enterprise will attain the best possible benefits in relation to their business. The AMP will be most effective if it addresses all the stages of the asset's life and makes use of synergies and opportunities to reduce costs and gain insight into plant condition.

The Asset Management Plan will be most effective if it is driven from senior management down and involves all the key maintenance stakeholders in its development. In operation it should be updated and reviewed at regular intervals to ensure that it is still relevant given changes in the business environment and plant.

There may be economic or practical constraints as to the scale and nature of the Asset Management Strategy. Nonetheless these compromises are best made in the context of a thorough understanding of the requirements of the particular plant. The availability of capable and experienced condition assessment practitioners will also have a significant bearing on what can be done, its frequency and cost.

The paper shows that there are a wide variety of methods available to gain insight into the condition of high voltage plant. What the tests do and their applicability for various plant types is covered. The value of Standards in relation to guaranteeing quality and the need to specify Standards and pass/fail criteria in contracts was emphasised. The benefits and difficulties associated with continuous on-line monitoring systems was covered and a view expressed that they are in general useful but that cost, ease of interpretation and their basic reliability were issues to be considered.

9 REFERENCES

[1] CIGRE Brochure 228 – Service aged insulation. Guidelines on managing the ageing process.

[2] Saha T., Review of Modern Diagnostic Techniques for Assessing Insulation Condition in Aged Transformers, IEEE Transactions on Dielectrics and Electrical Insulation Vol. 10, No. 5; October 2003.

[3] Christian J, Feser K, Procedures for Detecting Winding Displacements in Power Transformers by the Transfer Function Method, IEEE Transactions On Power Delivery, Vol. 19, No. 1, January 2004

10 BIBLIOGRAPHY

[1] Connell Wagner, HV Technicians Training Manual.

[2] IEEE (2000) Trial-Use Guide to the Measurement of Partial Discharges in Rotating Machinery, IEEE Power Engineering Society

[3] Lanham .C, The Need For High Voltage Testing In Predictive Maintenance Programs, http://www.reliabilityweb.com

MODELLING AND SIMULATION - A PROFITABLE TOOL FOR ALL PHASES OF THE LIFECYCLE

Mark D Mackenzie [a] and Ryan Briggs [b]

[a] IRC, 26 Colin Street, West Perth, Western Australia 6005, Australia.

[b] IRC,26 Colin Street, West Perth, Western Australia 6005, Australia.

Abstract: At any time during the life of an asset, decisions are made that impact its performance and therefore economic return. These decisions are made using a variety of tools and techniques which take into account a whole range of factors concerning the asset. Modelling and simulation is one tool that uses these factors to predict the behaviour of the asset. The benefit of modelling and simulation is that its flexibility allows it to assist in the decision making process through all stages of asset life. This paper uses case studies to illustrate the benefits of using modelling and simulation throughout the life of an asset.

Key Words: Modelling, Simulation, RAM

1 INTRODUCTION

At all stages in the lifecycle of an asset there is a risk that an inappropriate decision will be made which will have a negative impact on the availability of the asset. Using sound engineering theory, lessons learnt and in some cases engineering judgement, the risk can be reduced and the availability of the asset maximised.

Predicting asset availability is probabilistic, based on failure and maintenance rates associated with the systems, subsystems and elements within the asset. An appropriately built simulation model, populated with accurate data, can be used at every stage of the lifecycle to assist in the profitable design, operation and maintenance of the asset.

This paper discusses how modelling and simulation has been used at all stages of the lifecycle. Each stage is supported by a case study to illustrate how substantial cost savings can be made if modelling and simulation is used as a tool to assist in the decision making process

2 MODELLING AND SIMULATION

Simulation refers to a broad collection of methods and applications to mimic the behaviour of real systems, usually on a computer with appropriate software [1]. A set of assumptions are made in order to study how the system works. These assumptions, which usually take the form of mathematical or logical relationships, constitute a model that is used to try and gain some understanding of how the corresponding system behaves [2].

The modelling and simulation referred to in this paper uses the discrete event simulation software Optimise©. This type of modelling is ideally suited for observing the time based behaviour of a system. Individual elements of the system are represented as blocks. The relationship between the blocks is determined by the functional relationships that exist between the elements within a facility. The behaviour of the blocks is represented by the input parameters of the element. Typical input parameters include, but are not limited to Mean Time Between Failure (MTBF), Mean Time to Repair (MTTR), and failure effects or loss of capacity for each element. The input data is obtained from industry data sources or maintenance management systems. When data is not readily available an approximation can be obtained using a workshop attended by Subject Matter Experts (SMEs) who are familiar with the product and its operation. A graphical representation of the simulation process is shown in Figure 2-1.

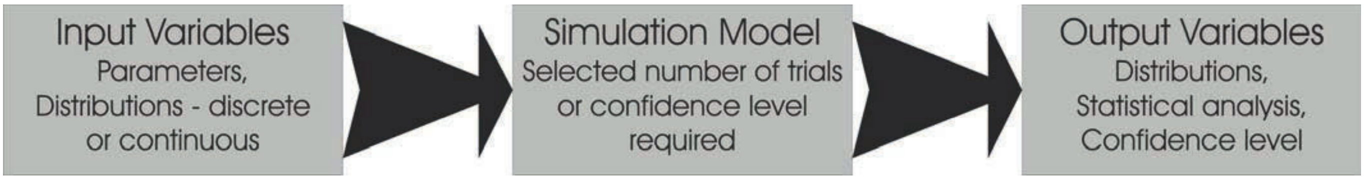


Figure 2-1 The Simulation Process [3]

The objective of modelling and simulation is to predict, with some degree of confidence, what the behaviour of the system (asset) will be under specific conditions. Each model is therefore built to answer a specific question. The answer being sought will determine how the model is built and what input parameters are required.

The reason modelling and simulation is so useful is that this "question" can be answered in a fraction of the time and cost of what it would take if the asset was allowed to operate for the actual time. Once the base model is built analysing different cases or "what if" scenarios is a relatively simple exercise.

Running the model consists of simulating a series of typical lifecycles. Each lifecycle is a realistic representation (based on the defined relationships) of the way the system may perform in operation. The model calculates the simulation results as averages over all the lifecycles. Simulation of multiple lifecycles also allows:

- Prediction of likely variation in performance (standard deviations); and
- Assessment of confidence levels in the predicted average performance.

Sufficient lifecycles are modelled in order that the results converge to statistically stable values.

3 ASSET LIFECYCLE

The life of an asset has been described as "Womb to Tomb" and "Cradle to Grave". What is really meant is that the lifecycle of the asset begins at its conception and finishes at its disposal. At all stages of the lifecycle someone is responsible for the asset and its cost effective management. There are many versions and theories on what constitutes the different phases of the asset lifecycle. The phases are not always easily defined and some phases may overlap. Examples of lifecycle phases include the Integrated Defense Acquisition, Technology, & Logistics Lifecycle Management Framework [4] model which has six lifecycle phases. The Systems Development Lifecycle (SDLC) [5] model has ten lifecycle phases. Blanchard [6] combines several phases to produce the following four lifecycle phases in which there are different opportunities to impact the design and cost effective operation;

The Design and Development phase includes the needs identification; options analysis; market research; maintenance concept; project initiation, and developmental testing. This phase is the most cost effective time to influence design as Figure 3-1 illustrates.

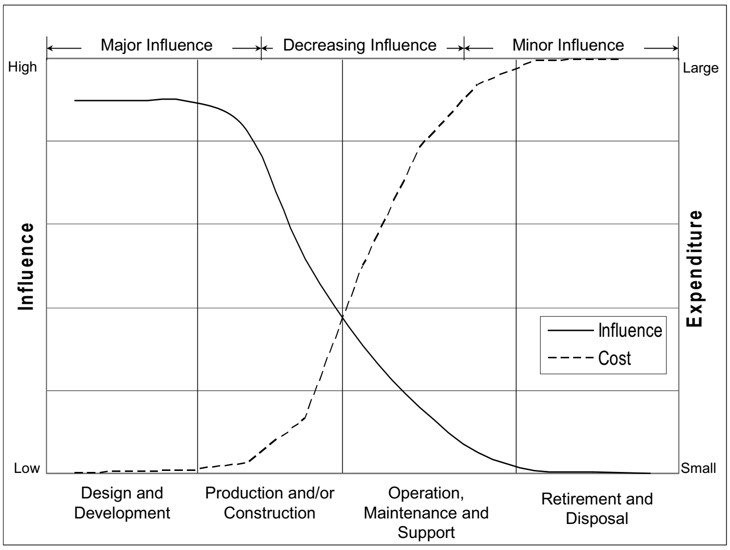


Figure 3-1 Cost and Impact during the Lifecycle Phases of a Asset [7]

The Production and/or Construction phase includes planning, contracting, system development, test and evaluation, technical reviews, maintenance planning, logistic support acquisition, and transition to the Operational Phase: Processes are presented for acquisition of each category of logistics support – support and test equipment, repair and overhaul services, software maintenance, facilities and renovations, drawings, manuals and software documentation, spare parts, personnel and technical training, engineering support.

The Operation, Maintenance & Support phase includes engineering support, operation and maintenance, problem identification and resolution, upgrades and modifications.

The Retirement & Disposal phase includes presenting scenarios and options, surveys and environmental audits, disposal planning and reporting.

Figure 3-1 illustrates the how the cost of influencing design increases and how the influence decreases as the asset life cycle proceeds.

4 COST EFFECTIVENESS OF MODELLING AND SIMULATION

The biggest cost involved in modelling and simulation is constructing the model to accurately represent the asset and populating it with appropriate data. The model may go through a number of iterations as it is refined to more appropriately represent the actual configuration, dependencies and relationships within the system. SMEs are an essential part of this process as their intimate knowledge of the system can be used to ensure the model is a true representation of the system being modelled. Obtaining the data to populate the model is often the single most difficult aspect of modelling the asset. However, once this has been completed the cost of making changes to design, production rates, maintenance philosophies, and life of the asset is easily determined.

A well constructed model that is developed during the design phase of the lifecycle can be developed and used throughout the entire lifecycle, up until and including the disposal phase. This is one reason why future use of the model is a cost effective tool as the majority of the cost is sunk. As updated operating and maintenance data becomes available the model can also be updated and simulations can be performed to determine if the asset is still likely to meet expectations. Making changes to the design, maintenance policy or trying to optimise availability can be done by simply modifying the model or updating the input parameters and running the required number of simulations.

5 DESIGN AND DEVELOPMENT PHASE

5.1 Overview

The design phase of the lifecycle is the time when the greatest cost savings can be made. As seen in Figure 3-1 if changes are required the most cost effective time to make them is early in the lifecycle.

When considering the design of any system or piece of equipment availability (capability) targets should be determined and specified during the design phase. As availability is a function of the asset's reliability and maintainability characteristics these characteristics must also be considered during the design phase. The availability targets will influence the design from the capacity and configuration selection to the maintenance and operating philosophy

If in a system, an item failure results in the total or partial loss of production then redundancy or increased capacity, respectively, may need to be built into the design. Consideration should also be given to the operational environment and the proposed maintenance program as they can have a significant impact on the overall availability, particularly when lengthy mobilisation periods occur prior to commencement of the maintenance activity.

Regardless of the type of equipment being analysed a preliminary design can be modelled and the results used to determine if the system availability is likely to meet the design and operational requirements. If the availability targets are not met after simulation then design changes can be implemented before construction begins. Updated models can be constructed and run in an iterative process until the desired targets are met. The following example illustrates a case study showing how simulation was used to influence the design to produce the most cost effective option.

5.2 Case Study [8]

5.2.1 Introduction

This case involved the creation of an integrated reliability based supply chain model to assist with selection of Liquefied Natural Gas (LNG) storage and fleet requirements. The supply chain model included all facilities and events that may impact on the performance of the LNG supply chain, including subsea production facilities, an offshore platform, an onshore LNG plant, LNG storage and port and shipping operations from the loading bay to various LNG receiving terminals.

A number of options with different combinations of LNG plant size, LNG tanker capacity and delivery locations were considered. The aim of this model was to determine, in terms of 'value' to the project, the 'optimal' LNG storage and fleet configuration for the option selected.

5.2.2 Method

As the project was in conceptual stage of design, project specific P&IDs and PFDs were not available for use. The model was constructed using design building blocks taken from the client's previous projects.

Modelling and simulation was used to investigate the following:

- the optimal shipping fleet configuration for given tanker sizes and delivery destinations;
- the optimal storage capacity at the LNG plant loading facility for each selected fleet configuration;
- the optimal LNG plant capacity; and
- the preferred LNG delivery split and delivery destinations.

Each option was assessed in terms of achieved LNG delivery and 'value' to the project. The 'value' parameter was determined by considering the revenue raised from additional LNG delivery and the costs associated with LNG tanker charter and additional LNG storage over the lifecycle of the project. The optimal solution was the best value result from the scenarios considered. The model is represented graphically in Figure 5-1.

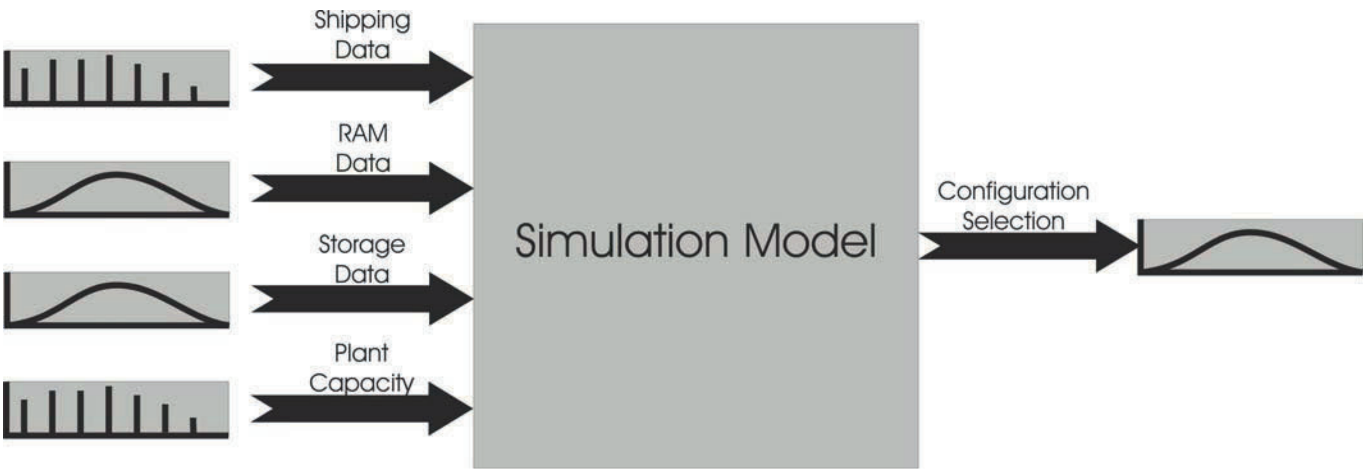


Figure 5-1 Plant and Shipping Model

5.2.3 Results

The study considered several configurations of LNG tanker size, delivery destination, RAM data, storage capacity, plant capacity and delivery split. For each arrangement, the optimal fleet configuration and LNG storage volume was determined. The results of a typical simulation run to determine the optimal LNG storage volume at the LNG plant loading facility for a given LNG plant capacity and a selected fleet configuration and LNG delivery split is shown in Figure 5-2.

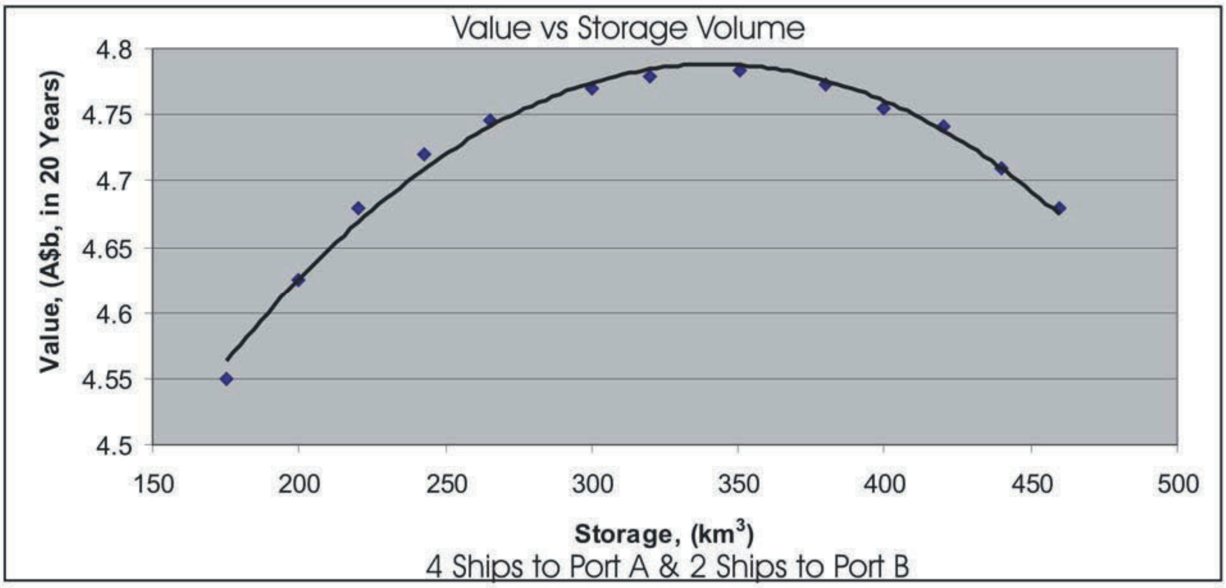


Figure 5-2 Value vs Storage Volume

In this case it can be seen that for the given fleet and plant configuration the "optimal" LNG storage volume is between 300,000 to 320,000m^3. This volume was determined to add the most value to the project I.e: the sale value of the additional LNG delivered with the extra storage volume offset the cost of additional storage. For storage volumes greater than 320,000m^3 the cost of extra storage is not compensated for by the sale of additional LNG.

5.2.4 Value of Modelling and Simulation

Modelling and simulation was used to support the decision making process during the Pre-FEED phase. Use of a reliability based integrated supply chain model enabled the selection of the fleet configuration and the LNG storage volume that provided the greatest value to the project.

Use of modelling tools in this stage of the design facilitated concept selection for the LNG plant, LNG storage and port and shipping operations. The final input parameters for the model can be used to form the basis for equipment reliability and performance specifications. It was essential to make these decisions as early as possible to allow the establishment of gas contracts with customers and for ship building to be undertaken within the lead times required by the vendors.

6 PRODUCTION AND CONSTRUCTION

6.1 Overview

Before production can begin the plant needs to be constructed and commissioned. During this stage the final design issues are resolved. This is last chance to influence design without serious cost consequences.

6.2 Case Study [9]

6.2.1 Introduction

In this case study a Hydrogen Sulphide Plant was to be upgraded to increase production. To support the design process a Failure Modes Effects and Criticality Analysis (FMECA) was requested to determine the impact of any potential failures on operations. As the design of the plant had been substantially determined two aims were identified: firstly to determine whether the current design could provide the required availability and production goal, and secondly to determine the impact on production of holding spares on site. Modelling and simulation was combined with the FMECA to help quantify the failure effect and to determine whether the plant had the capacity to meet the required production goals.

6.2.2 Method

A reliability block diagram, constructed from the P&IDs and PFDs, was reviewed by the engineering team to confirm the functional relationships and modified where required. A workshop, attended by operations, design engineers and maintenance personnel, was then conducted to determine the failure modes and their effects. The effect of holding spares on site for different equipment was a particular focus for the operations team. This was a significant issue as the effect on production due to maintenance activity delays associated with the lack of on-site spares had not been quantified.

The model was run for 1000 lifecycles of 20 years of operation. The input data for the model was provided from the workshop personnel. Two scenarios were analysed, one considered carrying no on-site spares the other considered carrying one spare of items that had maintenance delays while awaiting spares to arrive on-site. The model is represented diagrammatically in Figure 6-1.

Figure 6-1 FMECA Model

6.2.3 Results

The result of the analysis concluded that the proposed upgraded plant was capable of meeting the annual required production rates. The impact of holding spares on-site was an increase in the annual availability from 95.9% to 98% or an additional eight days of production per year.

Analysis of the sub-systems and elements within the system, to determine which were contributing the most to down time, was also conducted. The analysis revealed that the system contributing most significantly to the loss of availability was the Reactor and Quench Tower System. The typical contribution to loss of availability of each system is shown in Figure 6-2.

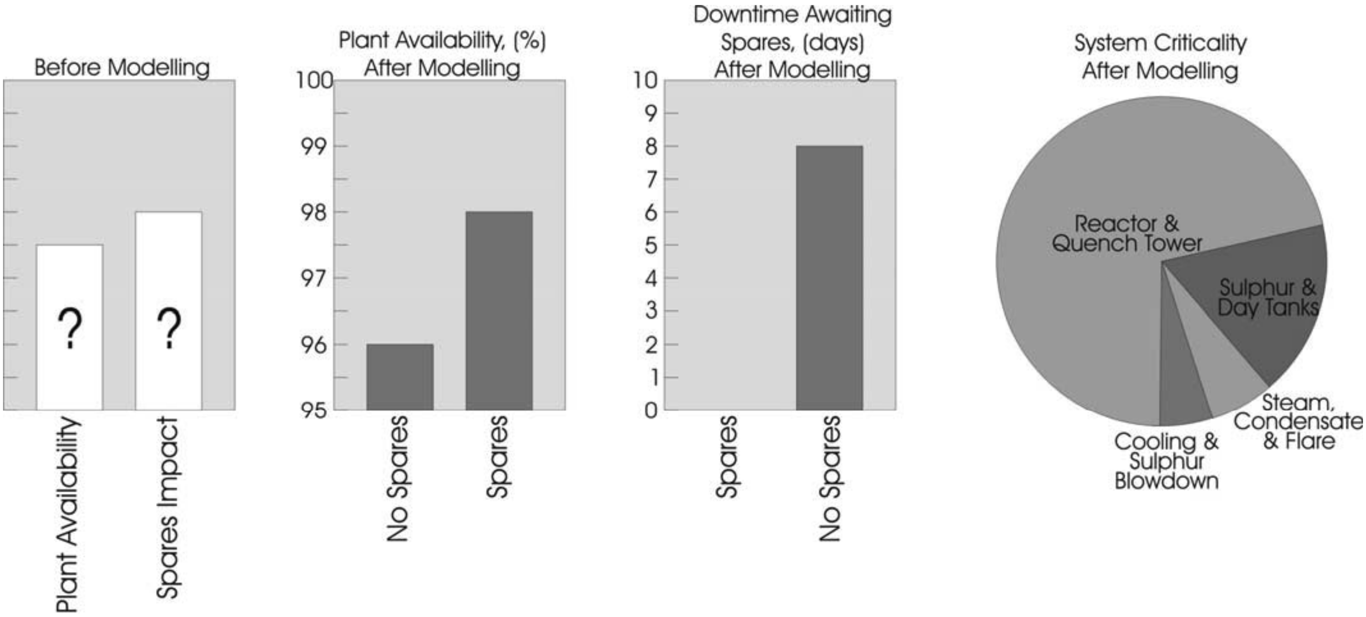


Figure 6-2 Simulation Results - FMECA

6.2.4 Value of Modelling and Simulation

Modelling and simulation was used to provide confidence that the proposed design would meet the required production requirements. The effects of holding spares on-site were quantified, allowing a cost benefit analysis to be performed to determine the value in holding spares on-site. The elements that contributed the most to the downtime were also identified to enable further engineering work to be done where required to further improve production.

In this case the modelling and simulation not only influenced the design but contributed to the operational and maintenance policy decisions. This was done with the cost of Lessons Learnt from the operation of the Plant.

7 OPERATIONAL PHASE

7.1 Overview

Operating and maintenance activities during the Operational Phase account for up to 75% of the lifecycle costs [10]. Modelling and simulation during this phase can be used to confirm design characteristics and assumptions. It is also used to de-bottleneck systems, subsystems and items that are having the greatest contribution to downtime.

A model built during the design phase can be updated during this phase to reflect the actual operating conditions. Any changes to the configuration, operations and maintenance activities to improve availability can be readily tested and refined using modelling and simulation before implementation of these strategies.

7.2 Case Study [11]

7.2.1 Introduction

During testing of the surface controlled subsurface safety valve (SCSSV) on a production well, the valve failed to fully seal. The well was shut-in awaiting assessment of the options available to enable production to recommence. SCSSVs are employed to shut down the well in case of a well blowout.

The client identified three options for the future of the well:

1. lock open the SCSSV and recommence production;
2. perform a well workover to replace the SCSSV and then recommence production; or
3. plug and abandon the well.

The aim of this model was to quantify the component risks for each option in monetary terms and estimate the total "opportunity cost" for each option by combining risk cost and actual cost.

7.2.2 Method

A fault tree was developed to represent the possible causes of a well blowout and the main risk assessment criteria was decided upon. Base data, including failure frequencies, Capital expense, operational expense and consequence costs for each risk that was not common to all options was then collated from a variety of industry sources. Using these data sources and risk assessment criterion, incident scenarios were developed and represented in an economic model to yield a monetary value. Capital expense, operational expense and risk costs were then combined to give an overall net present value (NPV) for each case, which allowed recommendations to be made.

7.2.3 Results

A risk based cost benefit analysis was developed to consider the three options available in terms of the lifecycle NPV costs. In each case an economic model was constructed to demonstrate the lifecycle costs and characteristics of each mode of operation over the remaining life of the well. Table 7.1 presents the summary of results for the three options.

Table 7.1 Summary of Results

Description	Option One	Option Two	Option Three
Environmental Risk	Negligible	Negligible	Negligible
Capital & Repair Costs	0	US$5,451,000	US$3,959,000
Production Costs	0	US$1,864,000	US$32,781,000
Risk Costs	US$2,600	US$9,100	0
NPV Cost	US$2,600	US$7,324,100	US$36,740,000
NPV Cost relative to Option One	0	US$7,321,500	US$36,737,400

7.2.4 Value of Modelling and Simulation

Modelling and simulation was used to support the decision to continue operations without a workover. The quantification of the risks in dollar terms helped demonstrate the value of option one. Other options or permutations of the existing options can readily be analysed using the base case that has been established.

Without modelling the three options a less financially rewarding option may have been selected. The expeditious use of modelling quantified the cost and risks (in dollar terms) of each option, allowing the production losses to be minimised.

8 LIFE EXTENSION OR DISPOSAL

8.1 Overview

There comes a time in the life of the asset when it is no longer profitable to continue operations with the existing plant. When profits are no longer high enough a choice to shut down operations completely or replace the entire plant may be needed. This phase of the lifecycle is the disposal phase where the residual value of the plant needs to be realised and all legal requirements met before divesting the asset management responsibilities.

Exactly when this time occurs for each asset is not easy to define. Life extensions, workovers, and economics changes could mean the asset is in operation for longer than originally predicted. Modelling and simulation can be used to determine the economics of continuing operations. It can also be used to determine if the entire asset or some small part needs to be upgraded to meet the production requirements.

8.2 Case Study [12]

8.2.1 Introduction

As part of an overall Iron Ore mining, processing and exporting expansion, the potential throughput of an existing crushing and screening plant needed to be determined. The crushing and screening plant consists of ten identical parallel screening units and four crushing units (one high capacity) in parallel with several material handling conveyors. A RAM analysis was conducted to assess the existing plant's availability at the proposed future production rate. This RAM analysis allowed the quantification of downtime and the identification of the major causes of downtime.

The aim of this model was to determine if the current equipment was suitable for the future throughput rate. If not, the quantification of RAM would determine whether the plant was suitable for upgrade or if a replacement was more appropriate. Popular opinion was that a replacement was required and $100 M was budgeted for the replacement.

8.2.2 Method

A workshop was held onsite to gather the required reliability parameters. The workshop participants included both site operators and project engineers. At the workshop, the equipment included within the crushing and screening plant was defined and divided into smaller systems for the purposes of gathering data, developing the RAM model and identifying areas of high losses and potential loss recovery.

The reliability parameters gathered in the workshop were used in conjunction with the project PFDs and operations and maintenance philosophies to develop the reliability block diagrams and a RAM model. The Optimise© RAM analysis software package was then used to simulate 1,000 lifecycles of the plant.

8.2.3 Results

Table 8.1 presents the summary of results of the RAM analysis for both the current and future throughputs. Increasing production increased the downtime and reduced the availability.

Table 8.1 Crushing and Screening Plant Performance Summary

	Performance Measure	
Case	Total Downtime, (hours per year)	Plant Availability, (%)
Current Throughput	1,200	86
Future Throughput	1,400	84

The modelling results showed a significant increase in downtime associated with production availability at the future throughput. The analysis of the subsystem criticality highlighted an increase in downtime associated with the plant screening and crusher systems. During current operations, the loss of a screening and crushing unit has minimal effect on achieving the desired throughput as there is some redundancy. However, at the future throughput, loss of either a screening unit or a crushing unit will have a significant impact on the downtime due to the lack of redundancy. Increasing the current throughput without upgrading the current screening or crushing units resulted in significant additional downtime annually. The results of the analysis are shown in Figure 8-1. If the screens and crushers are upgraded then the plant would be capable of meeting the production requirements. The cost to upgrade instead of replace was approximately $50M.

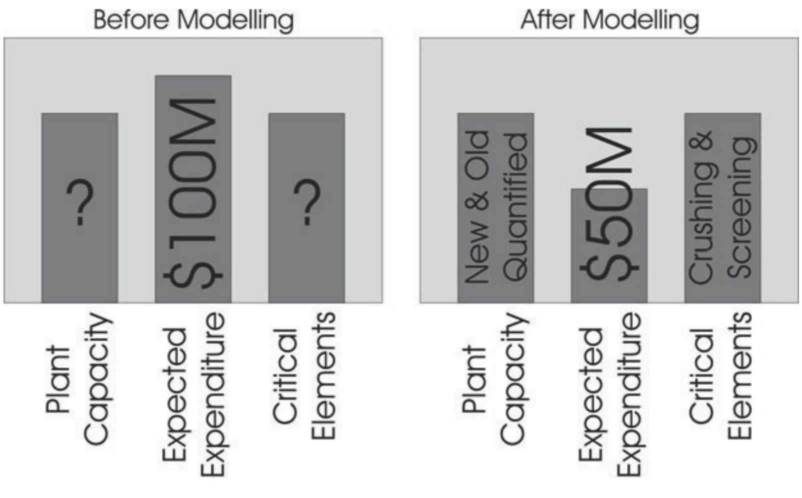


Figure 8-1 Simulation Results – Crushing Plant

8.2.4 Value of Modelling and Simulation

Modelling and simulation was used to support the decision to purchase new equipment or upgrade existing plant during the operational phase of the lifecycle. The quantification of the plant's potential future performance allowed the client to make an informed decision that the current equipment is suitable to be upgraded to meet future needs. The analysis identified the screens and crushing units to be the main systems which require upgrades.

Upgrading only the screening and crushing plant required less additional equipment than originally anticipated; the cost of the required upgrade would be approximately $50 million less than initially anticipated. Without modelling the plant an unnecessary upgrade may have resulted due to lack of understating of the plant. The model is now available if further upgrades are required and to validate their effectiveness.

9 DISCUSSION

The versatility of modelling an asset to gain some insight into how it will operate in the future and the benefit of knowing the effects of proposed changes make it an essential tool for any asset manager. As illustrated, using a few case studies, modelling and simulation can be used throughout the lifecycle.

Observations from industry indicate its use is more prevalent during the early lifecycle phases. Although the most benefit can be derived from its use at this stage it is rarely used in the later stages of the lifecycle.

Furthermore there appears to be little evidence that the models developed during the design phase are modified and used in the later phases of the asset's life. It appears there is a significant disconnect between the team responsible for designing the asset and the team responsible for operating and maintaining it. This is unfortunate as a significant leverage can be obtained by using a model that has already been paid for during the design phase. Using the model developed in the design phase to update RAM data and evaluate different maintenance and operational strategies is a simple exercise that can produce rapid results which can assist in the decision making process.

To assist in managing an asset throughout its entire lifecycle a greater awareness of the benefits of modelling and simulation needs to be highlighted. The communication between the design and operations teams needs to develop to allow the work undertaken in the design and development phase to be fully appreciated during the remainder of the life cycle.

10 CONCLUSION

The availability and performance of an asset is probabilistic with failures and maintenance activities following distributions specific to the each element. The dynamic interactions between the elements can be modelled using discrete event modelling and simulation. The results of the simulation can then be used to predict the performance of an asset before it has been built.

The model can then be used at all stages of the asset lifecycle to assist in decision making and provide the asset manager with a degree of confidence in the management strategy before it is implemented.

The most cost effective time to use modelling and simulation is during the design and development phase where it can assist in design and operational decisions. Full use should be made of the model by passing it to the operation team where it can be updated and maintained to reflect the latest configuration and conditions. Once the model has been updated it can be used to assist in decision making throughout the operation phase of the asset lifecycle. The model can also be used to assist in determining the most economical time for disposal.

11 REFERENCES

[1] Kelton WD, Sadowski RP, Sadowski DA. (1998) *Simulation with Arena*. New York, NY. McGraw-Hill

[2] Law AM, Kelton WD. (1991) *Simulation and Modeling Analysis*. New York, NY. McGraw-Hill

[3] Visser JK, Howes G *New Engineer Journal* August 2004

[4] http://www.dau.mil/pubs/IDA/IDA_04.aspx

[5] http://www.usdoj.gov/jmd/irm/lifecycle/ch1.htm#para1.2

[6] Blanchard BS (1998) *Logistics Engineering and Management* 5th Ed. Upper Saddle River, NJ. Prentice-Hall

[7] http://www.delphiaconsulting.com/s_architecture_reviews.htm

[8] RS-REP-05-128-001 Rev 0 dated 10 Jan 2006. (*LNG Supply Chain Analysis*)

[9] RS-REP-05-020-001 Rev 0 dated 29 Sep 2005. (*Failure Mode and Effects Analysis for the Hydrogen Sulphide Upgrade Project*)

[10] Blanchard BS (1998) *Logistics Engineering and Management* 5th Ed. Upper Saddle River, NJ. Prentice-Hall

[11] RS-REP-00-016-001 Rev 0 dated 18 May 2000. (*SCSSV study*)

[12] AO-REP-05-0124-001 Rev 0 dated 4 Aug 2005 (*Crushing and Screening Study*)

STATE WATER CORPORATION TOTAL ASSET MANAGEMENT PLAN

Barrie R[a] and Gonzalez N[b]

[a] *State Water Corporation, Manager Strategic Asset Services, 2-10 Wentworth St, Parramatta NSW 2150, Australia.*

[b] *State Water Corporation, Senior Asset Planning Engineer, 2-10 Wentworth St, Parramatta NSW 2150, Australia*

Abstract: The State Water Total Asset Management Plan (TAMP) documents the background, principles, processes, information and outcomes of its total asset management strategy. The TAMP provides a strategic context for present and future decision-making on asset management, encompassing risk, financial, environmental, heritage, levels of service for water delivery and engineering issues. Because of its holistic and strategic context, the TAMP has become a vital source of information and reference for all staff, customers and a wide range of stakeholders across government and commercial sectors. The TAMP has particular application in the deliberations of NSW Treasury and IPART when establishing full costs, efficient costs, sharing of costs between customers and government, and bulk water pricing. The TAMP has been a key link in State Water's relationships with the Customer Services Committees representing the interests of water users. State Water continues to maintain a rigorous program of continuous improvement on its 'Knowledge of Assets' and total asset management processes in general.

Key Words: asset, management, risk, financial, environmental, heritage, service, water, delivery, users

1 ORGANISATIONAL BACKGROUND AND CONTEXT

State Water Corporation (SWC) was established as a State Owned Corporation (SOC) on 1 July 2004, under the provisions of the State Owned Corporations Act 1989, and the State Water Corporation Act 2004. Prior to this, State Water was part of various Government Departments. SWC incorporates NSW's bulk water delivery functions outside of the areas of operation of the Sydney Catchment Authority, Sydney Water Corporation, Hunter Water Corporation, and of other water supply authorities.

SWC's principal objectives are 'to capture, store and release water in an efficient, effective, safe and financially responsible manner'. Its core delivery business provides services to about 6,200 regulated river customers. These services include asset operation and maintenance to deliver water from storages, metering and commercial services. The operation of these assets enables the delivery of about an average of 5,000 GL/year of water in the 14 regulated river systems (valleys), along 7,000 kms of river.

State Water's head office is located in Dubbo, however, the day-to-day administration of the delivery of bulk water services is carried out from four regional geographical Areas each of which with its own administration centre. The four Areas are North, Central, Coastal and South covering fourteen valleys.

State Water's organisational context, including regulatory environment, is depicted in Figure 1.

The NSW Government grants an Operating Licence, which authorises State Water to capture, store and release water; and to construct, maintain and operate water management works. The term of the current licence is three years and it commenced on 24 June 2005. IPART is the price regulator as well as the auditor of the Operating Licence on behalf of the Government. IPART is in the process of determining prices for the 2006/07-2008/09 period; and is yet to audit compliance with the conditions of the Operating Licence.

DNR is the natural resource manager and it provides State Water with a licence to access and use water by means of a Works Approval. This arrangement minimises clashes between water resource management and operation activities.

Treasury is both a shareholder and a part-provider of funding resources. IPART determines the magnitude of Treasury's funding contributions for the period of each determination. State Water is required to provide Treasury with return on and of assets.

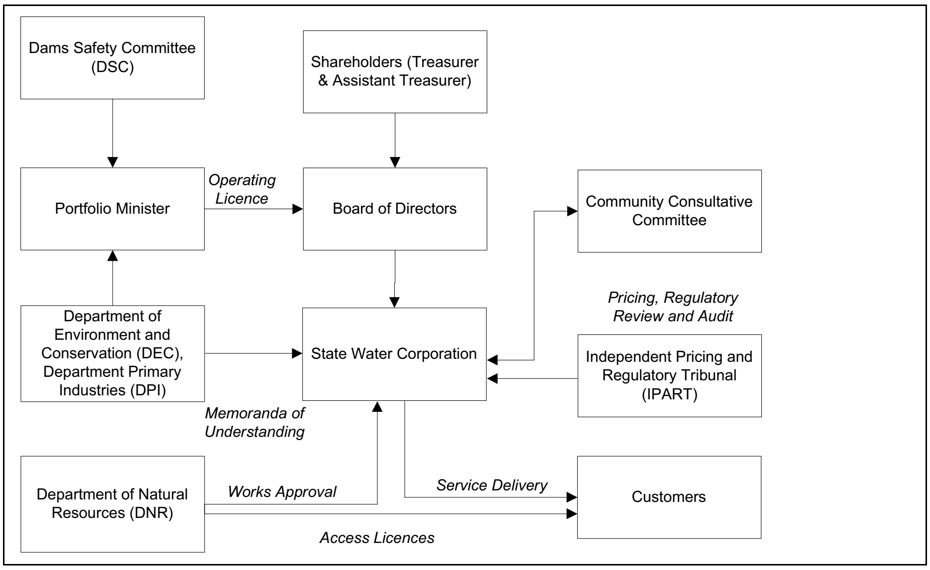


Figure 1 Organisational Context

2 ASSET BASE

As a bulk water service provider, State Water owns and operates an asset base worth $2.3b (Modern Engineering Equivalent Replacement Asset (2001 MEERA) value) or $300m (2004 regulatory asset base). This asset base comprises 20 large dams (all of them prescribed by the DSC), one small dam (prescribed by the DSC), approximately 300 weirs and regulators, 150 related river structures, 100 hydro graphic stations, approximately 500 buildings and numerous minor assets such as plant and equipment items. Additionally, it manages 11 small dams (three of them prescribed by the DSC) on behalf of NSW Department of Lands under a contractual arrangement.

The NSW Dams Safety Committee prescribes a dam if 'it has a potential for failure that could threaten downstream life, cause extensive property or environmental damage, or have a severe impact on the public welfare'.

3 ASSET MANAGEMENT DRIVERS

The key business drivers for the development and regular update of the TAMP include:

- Organisational Performance: Support corporate planning, implementation and allocation of resources, based on strategic asset management planning;

- Government policy: Premier's instruction on the principles of the NSW Government Strategic Management Framework, and the Treasury's Financial Management Framework including the government's Total Asset Management guidelines;

- Portfolio Risk Management: Provision of critical information for sound decision making. Information such as commercial and community risks, the development of risk management principles to define the duty of care and policy for the development of appropriate risk management strategies to remove, reduce or mitigate the risks; identify the linkages between assets, risks and insurance requirements;

- Compliance: Meet the statutory and regulatory requirements of NSW Dams Safety Committee, DEC, DPI and DNR;

- Corporate knowledge: Capturing long-term corporate knowledge will ensure informed decision making for planning, implementation and review. Avoid costs of repetition and of re-establishing lost information. Current information to support the preparation of budgets including CAPEX and OPEX;

- Public Good: Demonstrate to the Portfolio Minister who oversees public good, that community service obligations are delivered cost-effectively;

- Revenue: Meet the requirements of Treasury and IPART to develop plans of sufficient quality and detail as to enable the quantification of costs, and to meet cost setting guidelines;

- Cost Effectiveness: Demonstrate to customers, IPART, shareholders and community that asset management strategies are the most cost-effective, thus minimising costs for levels of service as defined in valley water sharing plans;

- Optimisation: The development of improved operations and maintenance strategies to optimise recurrent, renewal and capital expenditure;

- Benchmarking: Providing for rigorous measurement and benchmarking of all costs associated with the infrastructure, including operations and maintenance costs.

4 ASSET MANAGEMENT FRAMEWORK

As an infrastructure owner and bulk water service provider, State Water is chiefly an asset organisation. As such, asset management figures very predominantly within the corporate strategies. Figure 2 shows State Water's asset management framework.

Figure 3 further expands on the key asset management processes and asset management planning components of the framework and ties the detailed processes of the organisation into the overall framework. Both Figure 2 and

Figure 3 are colour-coded to facilitate visualisation of how the Asset Management Framework interrelates with Life Cycle Management Programs (LCMP).

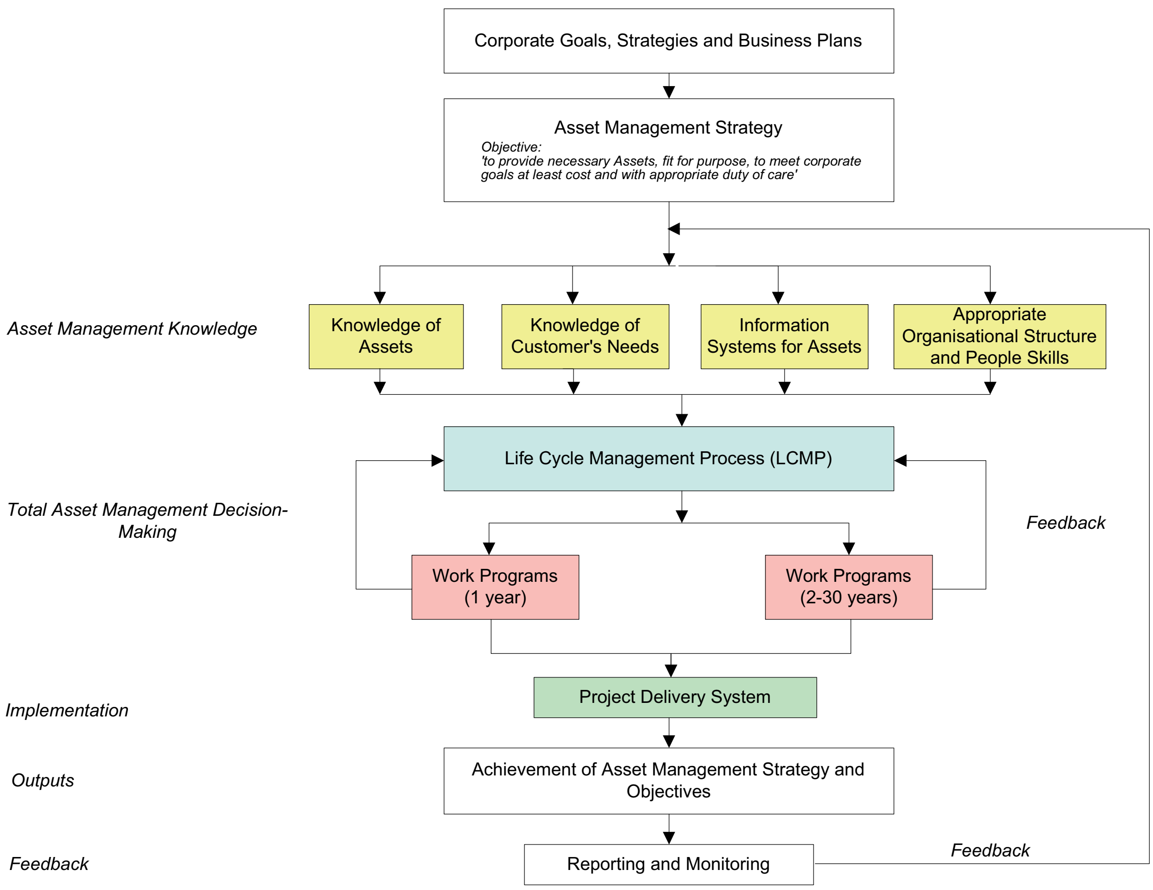


Figure 2 Asset Management Framework

The asset management framework is flexible and dynamic so as to minimise life cycle costs whilst delivering agreed levels of service to stakeholders. A very brief description of the components of the framework and how they integrate to deliver the Asset Management Strategy ensues.

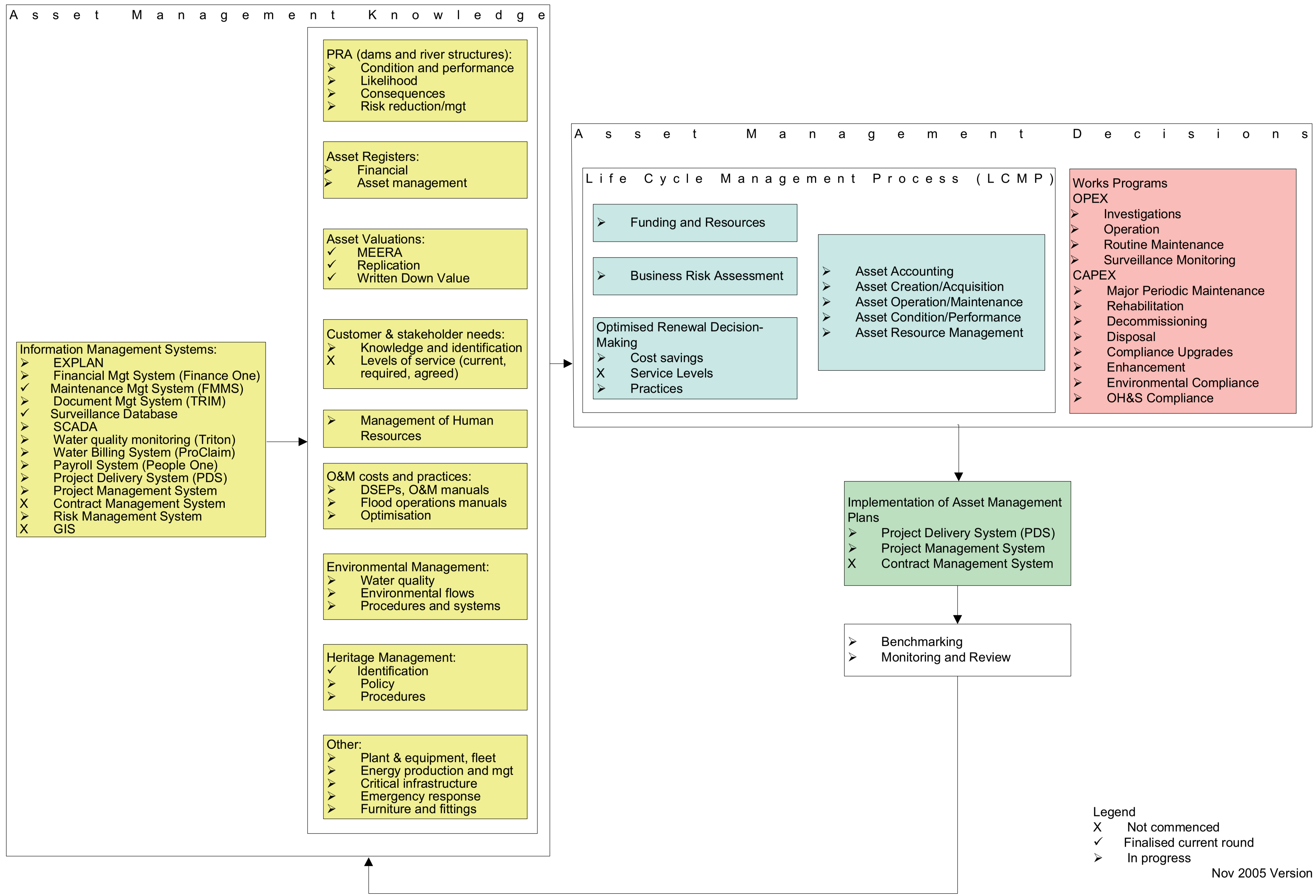


Figure 3 Life Cycle Management Programs

4.1 Knowledge of Assets

State Water is committed to continuously updating and improving the knowledge of its infrastructure because it realises its importance for asset management practices. Unless asset knowledge is regularly updated it will gradually deteriorate until it becomes unreliable and useless for asset management decision-making purposes.

Key components of various programs outlined in

Figure 3, which contribute to the improvement of asset knowledge, include:

- Condition assessments: State Water is in the process of developing a consistent quantitative rating system for asset condition for each of the various classes of assets. This program has two primary objectives, first to focus attention on components most likely to warrant immediate repair or further evaluation; second to monitor changes in the condition over time (and even at different locations). The latter is aimed at developing deterioration curves for predictive purposes;

- Maintenance and surveillance inspections: the frequency of these depends on the importance of individual structures. Some structures such as the large dams are subject to a strict program of surveillance inspection and maintenance audits according to requirements specified by the NSW Dams Safety Committee. Data collected from these inspections is recorded in registers and actions are tracked in subsequent inspections;

- Consequence assessments: Some assets, because of their importance, such as the large dams, warrant detailed consequence assessments involving population at risk (PAR), loss of life (LOL), direct and indirect damages, agricultural losses, intangible damages (environmental and cultural), etc. However, for the majority of the assets the consequence assessments are qualitative;

- Failure modes: the analysis of sequences or progression of events that may lead to failure is undertaken in detail for the large dams because of the complexity of the process and the expected consequences, should failure occur. State Water has developed a very good understanding of its portfolio of large dams, for which it carries out a thorough quantitative risk analysis, evaluation and control. A simplified process is being developed for the river structures as the risk drivers are more operational and the consequences are several orders of magnitude lower;

- Probabilities: Probabilities of failure, for different failure modes, are formally estimated in relation to the portfolio of large dams. For the remainder of the infrastructure base, probabilities are estimated on a qualitative base as part of risk analysis and assessment exercises;

- O&M practices and costs: operation and maintenance practices are recorded in O&M manuals that are regularly updated and which are available to users in hard copy at the different facilities or in electronic format through the document management system. Deviations from or additions to these procedures, particularly in relation to maintenance, are recorded in the Facilities Maintenance Management System (FMMS). The O&M manuals are updated regularly with these changes. Costs are captured both through the Financial Management System (Finance One) and through FMMS, at a more detailed level suitable for asset management purposes (e.g. by recording procedures followed, materials and plant used, etc.);

- Asset valuations: State Water conducts two sets of valuation. A Modern Engineering Equivalent Replacement Asset (MEERA) valuation and a replication valuation. The former it to satisfy regulatory requirements, the latter for asset management purposes, as the theoretical configuration of the asset may not reflect the existing asset layout. Valuations are used as proxy for value to stakeholders and are particularly useful for consistent benchmarking (internal and external) of performance (financial and asset management);

- Environment: State Water has developed, with input from relevant environmental agencies such as Department of Primary Industries (Fisheries), environmental protocols and checklists. These are designed to ensure compliance with environmental legislation and regulations;

- Heritage: State Water has completed a full assessment of heritage values for its entire portfolio of infrastructure and has set up a heritage register. It is now in the process of developing heritage management guidelines and conservation management plans for assets with identified heritage value.

4.2 Knowledge of Customers and Stakeholder Needs

State Water, as a State Owned Corporation, operates under the conditions of an Operating Licence, which formalises service delivery requirements to customers. IPART is charged with reviewing compliance with the conditions of the Operating Licence, which it is yet to do as the Operating Licence only came into effect on June 2004.

Some stakeholder requirements are clearly articulated through legislation including Occupational Health and Safety Act; State Water Corporation Act, Protection of the Environment Operations Act, Anti-Discrimination Act, State Owned Corporations Act, ICAC Act, Dam Safety Act and Protected Disclosures Act, amongst others. As such, they are not explicitly mentioned in the Operating Licence other than by reference.

A number of other less formal stakeholder requirements are negotiated in a collaborative manner through memoranda of understanding such as the one with the Department of Primary Industries for matters relating to fisheries.

There are eight valley-based Customer Services Committees (CSCs) that operate under agreed terms of reference. They are the primary means of consultation with customers on operational activities, pricing strategies, financial and business matters, and levels of service for water users within individual valleys. CSCs meet on a regular basis with senior management and the feedback they provide is duly incorporated in relevant processes. There is also a Customer Service Charter that guides the day-to-day interaction with over 6,200 regulated river customers.

State Water is committed to continuous improvement of service at the customer front. For this purpose it conducts customer satisfaction surveys and is in the process of developing an Issues Management Policy and accompanying computerised register, to manage customer/ public complaints. State Water also aims to be an employer of choice in the water industry and to that end it regularly conducts staff satisfaction surveys. The internal staff consultative committee has the responsibility to progress identified areas of improvement.

4.3 Information Management Systems

- State Water has been using a computerised Facilities Maintenance Management System (FMMS) for the last five years. This has assisted with planning and tracking progress to the extent that State Water is now quite confident that maintenance has reached a good standard;

- It has developed in-house a planning tool (EXPLAN) to assist with the IPART submission, forward capital investment planning and decision-making processes;

- It has recently acquired and is in the process of implementing a document management system (TRIM) and uploading information from previous systems into the new one;

- It has been using a procedural system to manage the delivery of projects (PDS) for the past three years. It is now investigating options to move into a computerised system for project management, audit and control with links to or integration with other corporate systems;

- It has a number of valley-based spreadsheet-type applications (CAIRO) for the day-to-day planning of water releases from storages and operation of river systems. It is now in the process of investigating the conversion of these into a central database application with live links with the water ordering system;

- The separation of State Water from its parent government department followed by its corporatisation in 2004 has forced it to acquire a number of computer systems that allow it to operate as a stand alone organisation. This process is still in progress and may continue for a few more years. The most important of these, is a package of three integrated applications for financial management (Finance One), water billing (Pro Claim) and human resource management (People One).

4.4 Human Resources

Prior to corporatisation, State Water had a staff base of around 240 employees including a small number of contractors. The corporatisation process brought with it a restructure exercise that had participation from relevant stakeholders. It reviewed organisational resources against the effects of statutory, operational and service delivery changes. The two most significant changes from the restructure were the creation of a new branch, risk management, and the establishment of a small group of engineering specialists. The former is aimed at addressing legislative requirements in a more rigorous manner from a business perspective. The latter seeks to improve the technical depth of the organisation, reduce reliance on consultants and account for the potential for expansion into external tendering. Overall staffing levels have increased across a number of organisational units, particularly in the areas of finance, information technology, human resource management and other corporate services. These changes are still in progress.

Another change derived from corporatisation has been the review and upgrade of a number of positions. This effort is aimed at attracting and retaining suitably skilled and experienced personnel. In addition, all segments of the organisation are locally formulating the skill requirements in detail to manage training needs accordingly, at local level.

A recent initiative is the set up of Technical Improvement Groups (TIGs). These are groups of skilled and dedicated staff across all areas of operations and abilities. They pursue excellence in work practices and technical knowledge in their specific areas of expertise and convey that expertise in the form of guidance on work practices and standards for the whole of the organisation.

Nine TIGs have been formally set up under terms of reference, they are:

- Lifting equipment and scaffolding;
- Corrosion control;
- Mechanical equipment;
- Electrical equipment;
- Major civil structures (dams, weirs and bridges);
- Emergency operation and stand-by equipment;
- Concrete and cement;
- Environmental engineering; and
- Contracts and contract management.

In 2005 State Water formed an internal storage management review committee to determine the most effective, efficient and economical framework to manage the risk of owning and operating large water storages and bulk water river assets. The objectives of the committee are to:

- Determine the most effective and efficient Area/ cluster/ site management structure for State Water Corporation storages that will enable it to meet regulatory and operational compliance requirements under both normal and exceptional circumstances;
- Recommend work practices that have higher value and will provide better short-term and long-term outcomes for the organisation, its staff and key stakeholders;
- Consider the impact on individual work sites, valleys, Areas and the whole of the State Water Corporation in making recommendations. (A variety of recommendations may result from differences in geographical separation, physical access conditions, site peculiarities or other factors.);
- Develop an understanding of the issues to be considered in an analysis of surveillance and operational work practices;
- Recommended changes to structures, systems, processes or work practices to allow increased operational efficiency with no reduction in customer service or compliance and no increase in risk to State Water Corporation;
- Increasing the quality of life for staff required to live on site for asset security; and
- Introduce higher value work practice through changed methods, systems, introducing technology or other means.

4.5 Asset Management Decisions and Life Cycle Management

The knowledge gathered through the seemly independent asset management processes, see

Figure 3 (e.g. asset valuations, risk, condition assessments, performance assessments, levels of service, O&M, etc.), is integrated together through the asset management framework, see Figure 2. All this information and recommendations from the individual processes comes together in the shape of works programs that take into account all these different drivers and which are prioritised according to alignment with the business objectives.

More specifically for large CAPEX works, State Water has formulated a 'capital investment and risk assessment framework' for decision-making. The framework is based on four principles formulated to ensure sound commercial outcomes for State Water, its shareholders and customers. These principles are:

- Complying with State Water's legislative and regulatory framework;
- Maximising State Water's net worth of investment by optimising social, environmental and economic outcomes;
- Managing risks within the business's risk tolerance levels whilst at the same time taking advantage of strengths and opportunities;
- Achieving consistency with ecologically sustainable principles in all investment decisions.

4.6 Implementation of the Total Asset Management Plan

Asset management plans are formulated on an ongoing basis and implemented annually through the budget process and in accordance with the rules of the project delivery system (PDS).

The PDS allows for the updating of information through the life of the project as it evolves. The template for project initiation and review includes business need, objectives, scope, constraints, stakeholder analysis, options comparison (including benefit cost analysis), risk management, communication plan, procurement options, etc. The PDS was introduced to:

- Minimise risk in the delivery of projects;
- Provide a system that is transparent and accountable;
- Ensure a uniform and consistent approach across the organisation;
- Improve communications and relationships through an understanding of the roles and responsibilities in delivering a project;
- Improve project risk management;
- Improve contract documentation and management;
- Ensure proper Environmental Impact Assessment of projects.

4.7 Benchmarking

State Water uses benchmarking as a means of reviewing its asset management practices, including decision-making, to continuously drive improvements. It does that by comparing resources and practices against those used in similar undertakings either within the organisation or outside. Some examples of benchmarking-type initiatives include:

- State Water conducted maintenance surveys over two consecutive years to assess its performance against industry standards. The results are very encouraging;
- It carried out internal benchmarking of maintenance costs and practices. The results highlighted a need for improvement in job costing practices;
- It participated in an international survey of dam spillway gate reliability the results of which were published to a selective audience in 2005;
- Initiated informal information sharing and benchmarking proposals with other organisations in the bulk water industry;
- Developed an internal technical exchange web page to promote information sharing;
- Identified a list of provisional key performance indicators and comparable organisations within Australia for benchmarking purposes.

4.8 Monitoring and Reporting of Performance

GHD has independently audited State Water's asset management practices and has conducted a number of gap analyses to compare its performance against best appropriate practice (BAP). Their reviews have shown that State Water has consistently improved its asset management activities and is able to bring a greater level of confidence to the planning and project tasks required for the delivery of bulk water.

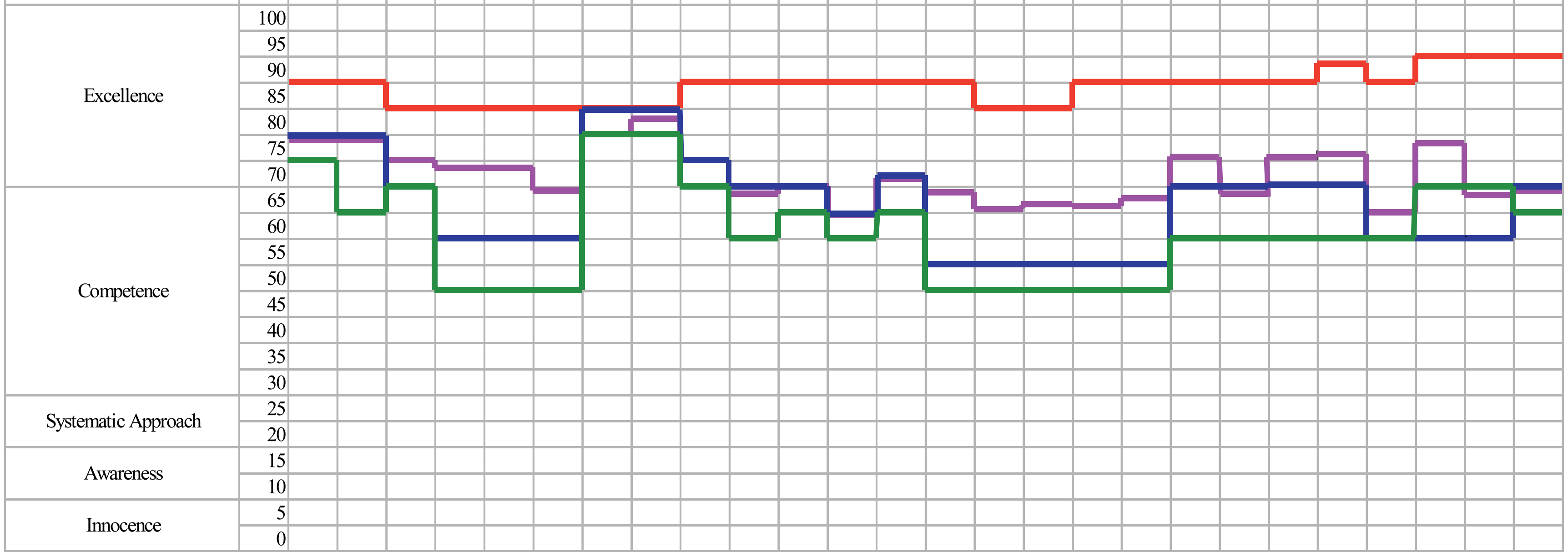
This conclusion is based on comparisons with earlier audits in 1997 and 2000, with key improvements observed in asset knowledge, risk management, project implementation and the use of information management systems. See Figure 4.

5 BENEFITS DERIVING FROM TAMP

The TAMP provides a strategic planning framework which allows:

- Identification and linking of business drivers to asset management planning;
- Providing the mechanism for balancing competing drivers for the whole portfolio;
- Consideration of an appropriate range of options, both on strategic and tactical matters, so that decisions can be optimised and expenditures minimised in the delivery of services and other outcomes;
- Management of individual and collective risks, both in the short term and the long term;
- Demonstration of duty of care through appropriate planning, maintenance and operational actions.
- Monitoring of performance and a mechanism for identifying improvements required and adjusting plans, processes and systems;
- Contributions to strategic asset management planning by all staff, as well as, staff planning their own activities in the context of the TAMP;
- Communication with regulators in considering their requirements;
- Production of key outputs, such as lifecycle management programs, capital works, operational and maintenance plans;
- Communication of the strategy to all external stakeholders, and an opportunity for stakeholders to test the relevance of asset management plans;
- Basis to secure funding for asset management.

Figure 4 Gap Chart



	Status	Overall Demand Analysis	Overall Knowledge	Overall Accounting and Costing	Overall Strategic Planning	Overall CAPEX Evaluation	Overall Business Risk	Overall Creation/Acquisition	Overall Rationalisation/Disposal	Overall Operations	Overall Maintenance	Overall Work/Resource Management	Overall Continuous Improvement	Overall LCM Processes and Practices
Reference Number		1.01	1.02	1.03	1.04	1.05	1.06	1.07	1.08	1.09	1.10	1.11	1.12	1
Best Appropriate Practice	87.8	90	90	85	85	85	85	85	85	90	90	90	90	88
Gap Analysis 2004	69.2	78	78	75	68	67	64	80	83	75	68	70	64	72
Gap Analysis 2000	64.7	80	80	70	60	60	60	85	85	75	70	70	65	72
Gap Analysis 1997	61.5	75	65	70	50	50	50	80	80	70	60	65	60	65

	Overall Primary Systems	Overall Secondary Systems	Overall Tertiary Systems	Overall Systems Issues	Overall Rating- AMIS	Overall Primary Data	Overall Secondary Data	Overall Tertiary Data	Overall Data Rating	Overall Commercial Tactics	Overall Organisational	Overall People Issues	Overall TAMP Rating
Reference Number	2.01	2.02	2.03	2.04	2	3.01	3.02	3.03	3	4	5	6	7
Best Appropriate Practice	90	85	85	90	88	90	90	85	88	85	90	90	90
Gap Analysis 2004	68	61	62	61	63	76	68	71	72	65	78	68	69
Gap Analysis 2000	55	55	55	55	55	70	70	70	70	60	60	60	65
Gap Analysis 1997	50	50	50	50	50	60	60	60	60	60	70	70	65

6 ASSET MANAGEMENT CHALLENGES FOR THE FUTURE

State Water's approach to asset management has been, and is expected to continue to be, subject to continuous change and adaptation given the state of the water industry and new developments in technology. Asset management is expected to undergo further changes to accommodate the operation of the organisation under commercial considerations. Decisions will be scrutinised more rigorously to demonstrate quadruple bottom line objectives and a level of water charges that enables the organisation to be self-sufficient.

Challenges for the future include:

- Modelling of asset condition deterioration based on history of condition rating assessments collected over time;
- Continuous refinement of useful and residual asset lives through condition and performance assessment programs. In particular the blend of these two concepts, physical condition and performance, needs to be clearly and consistently defined to assess how it affects the ability of an asset to meet its agreed level of service;
- Predictive modelling and optimisation (including sensitivity analyses) of asset replacement/ rehabilitation/ maintenance strategies and impacts on levels of service;
- Adaptation to international financial reporting requirements. This will not only force some changes in the way that the organisation accounts for costs but also have effects on who pays for what;
- Financial modelling of capital works program (based on a number of forecast cost profiles) aimed at formulating financial strategies to fund renewals. Alternatively, financial constraints can be set in the model to develop asset renewal budgets.

STATE WATER PORTFOLIO RISK ANALYSIS

Brian Cooper[a] and Robecca Gilham[b]

[a] *Dams & Civil, Department of Commerce, Rawson Place, Sydney New South Wales 2000, Australia*

[b] *Strategic Asset Services, State Water Corporation, Wentworth Street, Parramatta New South Wales 2150, Australia*

Abstract: State Water Corporation (SWC) is a bulk water supply business which manages a portfolio of large dam structures prescribed under the Dams Safety Act (1978). The Act sets out deterministic criteria to ensure that dam owners apply adequate consideration to dam safety for structures under their control. With limited funding and what may seem to be endless upgrading programs to keep pace with revised hydrology, in recent years dam safety management has moved away from application of traditional deterministic criteria, toward a risk based approach. Risk assessment provides the means to trade off the varying needs of structures in a portfolio, enabling owners to prioritise limited funding resources to areas likely to have the greatest impact and benefit. The process engaged to analyse risk in SWC's portfolio, and further to develop its capital works upgrading program will be the focus of the following discussion.

Keywords: Bulk water, dam safety, hydrology, risk analysis.

1 ABOUT STATE WATER CORPORATION & ITS REGULATORY ENVIRONMENT

State Water Corporation is wholly owned by the Government of New South Wales. The Corporation's principal objectives are *'to capture, store and release water in an efficient, effective, safe and financially responsible manner'* – for use by customers and the community – and for preservation of riverine ecosystem health. Water allocation to meet competing demands is governed by a Statutory Water Sharing Plan. State Water operates within the plan and its revenue is effectively capped by the plan limit.

As an owner/manager of large dams SWC is regulated by the Dams Safety Committee (DSC), a NSW government statutory authority created under the Dams Safety Act. Its role (among other things) is to *"formulate measures to ensure the safety of dams"* and to *"maintain a surveillance of prescribed dams"* (Dams Safety Act 1978). A "prescribed dam" is one listed in Schedule 1 of the Act.

In addition to the DSC, SWC is guided by the Australian National Committee on Large Dams (ANCOLD). ANCOLD's goal is to encourage improvement in the planning, design, construction and operation of large dams in Australia, and to ensure that dam owners have access to world best practice through the skills of Australian professionals.

ANCOLD has published the Guidelines on Risk Assessment (October, 2003), which has shown a significant move away from the deterministic approach to dam safety management. This Guideline has seen ANCOLD embrace the concept of risk, and has been a major source of guidance for SWC Portfolio Risk Analysis.

2 INTRODUCTION

The portfolio of dams owned by SWC presents it with a certain exposure to risk. The risk presented by some dams will be greater than by others. Some dams will satisfy deterministic standards for dam design while others will not, and will require remedial works to bring them up to standard. Of this latter group of dams, some will cause greater loss of life (LOL) and economic damage if they were to fail than others. It is apparent then, that there needs to be some means to rank the dams regarding their likelihood of failure and the associated consequences of failure. The Portfolio Risk Analysis (PRA) provides these means. The PRA is currently in the third stage of an extended risk management exercise that is briefly outlined below.

The first stage of risk assessment in its current form was the Screening Level Risk Analysis (SLRA). The SLRA was completed in 1999, and was a first pass at assessing the risk for 16 major dams (Hume Dam, Menindee Lakes and Fish River structures were not included) and 14 minor (ex-CaLM) dams that were under the control of SWC at this time. The SLRA developed probabilities of failure using event trees, however only made a rough estimate of LOL.

Following on from the SLRA, was the PRA (2002). This study developed the probabilities of failure (via event trees) from those estimated in the SLRA, and assessed the consequences of dam failure including LOL, economic damages, and environmental/cultural damages. A separate consequence assessment report was prepared, by consultants Sinclair Knight Merz, for each major and minor dam in the portfolio.

The current revision of the PRA is focussed on updating information that has become out of date, extending the study to include all prescribed dams managed by SWC, as well as developing the usefulness of PRA (2002) as a tool to guide SWC's program of capital works. The structures included in this revision of the PRA are listed in Table 2-1 below.

Table 2-1 Prescribed Dams in the Current PRA Revision

Dam Name
Blowering Dam
Brogo Dam
Burrinjuck Dam
Burrendong Dam
Chaffey Dam
Carcoar Dam
Copeton Dam
Glenbawn Dam
Glennies Creek Dam
Hume Dam
Keepit Dam
Lostock Dam
Menindee Lakes Storage's
Oberon Dam
Pindari Dam
Rydal Dam
Split Rock Dam
Toonumbar Dam
Windamere Dam
Wyangala Dam

3 ASSET MANAGEMENT FRAMEWORK FOR DAM SAFETY

The Role of the DSC

The DSC has statutory functions under the Dams Safety Act 1978 and the Mining Act 1992. Its main objective is to ensure that all prescribed dams in NSW are in such a condition as to not pose an unacceptable danger to downstream residents and property, or to adversely affect the public welfare and environment. This is achieved by requiring all dam owners to arrange for:

- Regular monitoring and surveillance of their dams;
- Ongoing assessment of their behaviour on the basis of monitoring and surveillance information;
- Regular review of the compliance of their dams with current standards;
- Review of all such information and assessments by experienced personnel;

The DSC has an ongoing watchdog role to oversight the safety of the State's prescribed dams and to prevent significant inadvertent or uncontrolled loss of their stored waters. In this context safety means that the dam complies with the Committee's current requirements and conforms to current accepted national and international practices.

The Committee requires owners of prescribed dams with known, or assessed potential, deficiencies to formulate appropriate remedial action programs and to implement appropriate surveillance and emergency plans. These actions are to be undertaken with a degree of urgency that is dependent on the extent of the deficiency and the scale of the consequences in the event of dam failure.

3.1 The Role of ANCOLD

ANCOLD is the Australian national committee of the International Commission on Large Dams (ICOLD), a non-government organisation established in 1928. As part of ICOLD, ANCOLD guidelines are generally representative of current accepted national and international practices. ANCOLD is an incorporated voluntary association of organisations and individual professionals with an interest in dams in Australia. Members include local, state and federal agencies, dam owners and operators, contractors, consultants and academics. Many disciplines are represented including planners, environmental scientists, engineers, hydrologists, geologists, social scientists, economists and legal practitioners.

ANCOLD's goal is to encourage improvement in the planning, design, construction and operation of large dams in Australia, and to ensure that dam owners have access to world best practice through the skills of Australian professionals. ANCOLD pursues this goal through a number of activities, however the publication of Guidelines setting national standards in many aspects of dams and their environs, are summaries of asset management activities which together constitute best practice asset management of dam structures. ANCOLD guidelines cover topics including, but not limited to the following:

- Risk Assessment
- Dam Safety Management
- The Environmental Management of Dams
- Assessment of the Consequences of Dam Failure
- Selection of Acceptable Flood Capacity for Dams
- Design of Dams for Earthquake
- Strengthening and Raising Concrete Gravity Dams
- Design Criteria for Concrete Gravity Dams
- Concrete Faced Rockfill Dams

3.2 Asset Management within the Regulatory Environment

In terms of dam safety, SWC has a statutory obligation to comply with the requirements of the DSC, and a professional responsibility to comply with ANCOLD guidelines and hence demonstrate best practice and due diligence.

Traditionally asset management in dam safety is composed of a number of areas, including design, operations and maintenance, surveillance and risk management and these are discussed in the topics of the ANCOLD guidelines listed above. Risk assessment is the most recent addition to the asset management tools, and it is important to remember that this provides one input into an asset management process.

ANCOLD is currently leading change in Australia, away from deterministic standards. The DSC is undertaking a review of the regulatory policy framework for dam safety. This revised policy, if implemented, will represent a large shift away from purely deterministic statutory dam safety standards, and move to focus on an integrated dam risk management process which includes risk analysis, risk evaluation, risk communication and risk control.

4 WHAT IS RISK?

4.1 Risk Analysis, Evaluation, Assessment and Management

There is often confusion between the terms Risk Analysis, Risk Evaluation and Risk Assessment. To clarify this difference, ANCOLD definitions for each follow:

4.1.1 Risk Analysis:

The use of available information to estimate the risk to individuals or populations, property or the environment from hazards. Risk analyses generally contain the following steps: scope definition, hazard identification, and risk estimation.

4.1.2 Risk Evaluation:

The process of examining and judging the significance of risk. The risk evaluation stage is the point at which values (societal, regulatory, legal and owner’s) and judgements enter the decision process, explicitly or implicitly, by including consideration of the importance of the estimated risks and the associated social, environmental, economic, and other consequences, in order to evaluate a range of alternatives for managing the risks.

4.1.3 Risk Assessment:

The process of reaching a decision recommendation on whether existing risks are tolerable and present risk control measures are adequate, and if not, whether alternative risk control measures are justified or will be implemented. Risk assessment incorporates, as inputs, the outputs from the risk analysis and risk evaluation phases.

Together risk analysis and risk evaluation, represent risk assessment. Risks are first analysed and then compared to tolerability criteria, and hence the risk is evaluated. The ANCOLD (2003) guidelines are directed to the practical application of risk assessment, with a primary focus on risk to life. The guidelines therefore provide guidance on tolerability criteria within the dam safety framework.

The SWC PRA is generally only concerned with risk analysis in terms of dam safety, with risk evaluation and assessment only really being introduced with respect to risk to life. ANCOLD provides guidance in the tolerability of societal risk, through the ANCOLD Societal Risk Curve and individual risk limits. However economic, environmental and cultural risks remain a matter for a business to decide upon a tolerable level. SWC has not considered the tolerability of economic, environmental and cultural risk at the present time. The outputs from the present process to revise and develop the PRA (2002) will remain a risk analysis with risk evaluation in terms of life safety risks.

Risk management is the end objective – risk assessment is a means to that end. In SWC, we have developed a risk analysis process, and assessed risk to life with respect to the ANCOLD guidelines. This process is the subject of this paper. The tolerability of risk in terms of economic, environmental and cultural risk remains yet to be addressed.

4.2 General

To quantify risk, the basic equation of Risk = Probability x Consequences was applied. The total risk is then the sum of all risks for the number of events considered, as is represented as:

$$\sum_{i=1}^{n} \{P(fail)_i * (csq(fail)_i\}$$

Where:

P(fail)i	is the probability of failure path i on the event trees
csq(fail)i	is the consequence of failure for failure path i (loss of life)
n	is the number of failure paths considered

4.3 Probabilities of Failure

4.3.1 General

Risk has been treated on a dam by dam basis, therefore to determine the probability of failure for a single dam, the structure is first broken down into its components (these may include the main embankment, spillway, reservoir and any subsidiary dams). Failure of these individual components is then considered separately.

The process of assessing the probability of failure of each dam component under various hazards follows a similar path. Hazards that have been included in the PRA assessment include earthquake, flooding, and sunny day (normal) events. In PRA (2002) piping was considered as a separate failure mode, however current practice is to assess piping following from one of the above hazards, using an event tree approach. The overall failure of the dam may result from a combination of the above hazards and the dam components not performing in a satisfactory manner.

4.3.2 Event Trees

Event trees have been developed to separate the dam components into events that are mutually exclusive, and therefore may be summed to calculate an overall probability of failure for the component, or the dam system. This overall probability of failure would therefore consist of earthquake, flooding, and sunny day events, relating to the various dam components. A sample even tree is shown in Figure 4-1 below.

Looking at Figure 4-1 we can see that the tree starts with a single node ‘F*R’, that is known as the initiating event. ‘F’ tells us that the type of initiating event is flood, and the characters following describe the particular flood partition. The probability of a particular flood partition occurring is the probability of the particular initiating event. Partitioning of annual exceedence curves within event types facilitates a more refined calculation, as it allows consequences relevant to the magnitude of each event being considered to be applied. Without partitioning, only one set of consequences would be applied to a probability of failure, which may, for example, overstate the risk. The initiating event is then followed by a number of branches (four in the example below), each with probabilities attached. The probabilities following the initiating event are known as conditional probabilities.

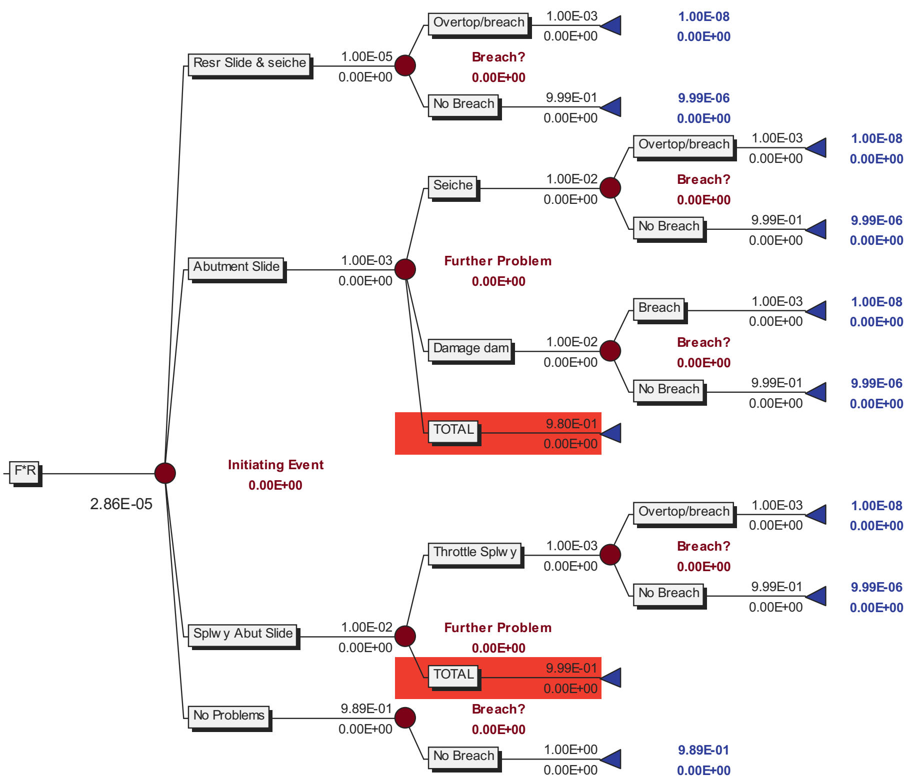


Figure 4-1 Sample Event Tree for Reservoir Failure

4.3.3 Conditional Probabilities

Conditional probability is the chance of an event occurring given that another event has already occurred. As with the SLRA, the conditional probabilities were estimated generally using expert judgement, on the basis of Table 4-1 below, supported by information contained in original design drawings, design reports, geotechnical investigations and design reviews, where information is available.

Table 4-1 Conditional Probabilities Mapping Scheme after Barneich et al (1996)

Description of Condition or Event	Order of Magnitude Annual Probability
Occurrence is virtually certain	1
Occurrances of the condition or event are observed in the database	10^{-1}
Occurrence of the condition or event is not observed, or is observed in one isolated instance, in the available database; several potential failure scenarios can be identified.	10^{-2}
The occurrence of the condition or event is not observed in the available database. It is difficult to think about any plausible failure scenario; however, a single scenario could be identified after considerable effort.	10^{-3}
The condition or event has not been observed, and no plausible scenario could be identified, even after considerable effort	10^{-4}

Reprinted from: ANCOLD Guidelines for Risk Assessment, 2003

The amount of data being analysed in the PRA can be appreciated if we consider the following:

- 20 prescribed structures make up the portfolio
- most structure have event trees for at least the main embankment, spillway and reservoir, and many have more than one subsidiary embankment

- all structures are analysed under the hazards of earthquake, flood and sunny day
- initiating events are often sub-divided further into partitions, which represent a level of consequences of failure. Each initiating event then required a set of consequences commensurate with the magnitude of the initiating event to be developed.

4.3.4 Joint Probability

Joint probability of failure is the chance of breach in any given year owing to a specified initiating event and set of conditions. Joint probabilities are determined by multiplying along each of the branches of an event tree. In the PRA (2002), the sum of the joint probabilities for all the initiating events on a particular event tree has been taken to be the probability of failure for that component of the system under the specified loading condition.

Care should be taken to determine the overall probability of failure as a result of all failure paths as when extreme events are estimated there may be a number of potential failure paths, all of significant conditional probabilities. In this case it is misleading to simply sum the probabilities of each branch, as the result may exceed unity, which is impossible. The uni-modal bounds theorem addresses this problem.

PRA (2002) has generally resolved this matter, through application of de Morgan's Rule between individual components within each initiating event. From de Morgan's rule, the estimated upper bound conditional probability (before application of the initiating event) is equal to one minus the product of the complements, or expressed as:

$$P_{UB} = 1 - (1 - P_1).(1 - P_2)\ldots\ldots(1 - P_n)$$

Where:

P_{UB}	= the estimated upper bound conditional probability of failure
P_1 to P_n	= the estimates of the several individual mode conditional probabilities of failure.

Technically, this rule should be applied between components of the dam system, as well as between failure mode types, before application of the probability of the initiating event. The influence of failure modes/components on one another, however, is generally minimal, since each conditional probability is a relatively small number, ie <0.01.

4.4 Consequences of Failure

4.4.1 Introduction

The consequences as a result of failure of each of the dams in the PRA (2002), were examined in a study carried out by Sinclair Knight Merz (SKM). The SKM study produced a report for each dam and in general analysed consequences of failure for failure (and no failure) modes including: Sunny Day Failure; Probable Maximum Precipitation Design Flood (PMPDF) with fail; Probable Maximum Precipitation Design Flood without fail; Dam Crest Flood (DCF) with fail; and Dam Crest Flood without fail.

The incremental consequences were estimated by subtracting the results of the no fail cases from those of the fail cases. The ANCOLD Guidelines on Risk Assessment (October 2003) now recognise the Graham (1999) method (used to derive LOL in dam failure cases) should not be applied to the without failure case. The consequences for the Sunny Day Failure represent incremental values directly because in this case the consequences of no failure are considered zero.

For dams where current dambreak studies and hence inundation plans were available, these were utilised to derive consequences. Otherwise, numerical hydraulic modelling was carried out to simulate the effects of flooding, and flood inundation maps were prepared. Three main groupings of consequences were derived including LOL, economic loss and environmental and cultural losses. These broad groups are described in greater detail below.

4.4.2 Loss of Life

The Graham (1999) method divides the fatality rate (loss of life as a proportion of a population at risk due to dam failure) into 15 different possible combinations; 3 flood severity categories, 3 warning time categories, and 2 flood severity understanding categories. Combinations of these categories assign a recommended fatality rate and a recommended range of fatality rates. It is therefore possible to not only predict a best estimate of the fatality rate from Graham's recommendation and site specific information, but also present the possible range. A range of LOL was presented in the PRA (2002).

4.4.3 Economic Consequences

The economic consequences from flood damage are both tangible and intangible. The tangible losses are those arising from destruction and losses which can be measured in terms of their market values. They comprise direct and indirect damages. Direct damage is that damage to property that is a direct consequence of flooding and includes:

- destruction and damage to homes, commercial/industrial structures, building contents, goods and equipment
- public building including schools, hospitals, libraries
- destruction and damage to infrastructure including roads, bridges, power lines, water supply pipelines, sewage treatment plants and other utilities
- vehicles
- mines and power stations
- farm assets
- agricultural, pastoral and stock losses
- agricultural losses due to loss of irrigation water.

- Indirect damages are those which flow on from the direct damages and include:
- cost of temporary accommodation
- cost of emergency services
- cost of traffic disruption
- loss of retail trade
- loss of industrial production
- temporary costs for provision of water and sewerage services
- lost agricultural and pastoral production in periods after flooding
- subsequent health costs that can be costed and attributed to the flooding.

Intangible damages are those damages that do not have a market value that can be easily measured. Factors such as anxiety and ill health cost the community but the costs are not readily measured. The consequences assessed in this study only quantify the tangible losses.

The damage costs presented are generally economic costs rather than financial costs. Economic costs are those costs measured in terms of consumption of resources from a community aspect. In this case, the community is the whole of Australia. Financial losses focus on the losses suffered by a single individual or group of individuals, businesses or government agency.

4.4.4 Environmental and Cultural Consequences

Dam failure is likely to cause significant environmental damage and cultural loss. A qualitative assessment involved classifying the ecological and cultural value of the following sensitive areas:

Environmental Consequences:

- Riverine Systems
- Wetlands
- Terrestrial Species
- Threatened Species
- Effect of Major Point Pollution

Cultural Consequences:

- European Heritage Sites
- Aboriginal and Heritage Sites
- Parks and Reserves

Impacts were ranked from 1 to 5 with negligible effect at th**e low end of the scale and total destruction at the high end of the scale. Intermediate impacts were minor, moderate and major.**

4.5 Calculating Risk

4.5.1 Risk to Life

Risk to life is comprised of both societal and individual risk. The ANCOLD definition of life safety risks is outlined below:

Individual risk:

The increment of risk imposed on a particular individual by the existence of a hazardous facility. This increment of risk is in addition to the background risk to life, which the person would live with on a daily basis if the facility did not exist.

Societal Risk:

Also referred to as societal concerns. The risk of widespread or large scale detriment from the realisation of a defined hazard, the implication being that the consequence would be on such a scale as to provoke a socio/political political response, and/or that the risk (that is, the chance combined with the consequences) provokes public discussion and is effectively regulated by society as a whole through its political processes and regulatory mechanisms. Such large risks are typically unevenly distributed, as there are attendant benefits. Thus the construction of a dam may represent a risk to those close by and a benefit to those further off, similarly a process may harm some future generation more than the present one.

To assess the tolerability of societal risk, SWC made reference to the ANCOLD Societal Risk Curve. On this curve, the cumulative frequency of the probability of failure of each initiating event is plotted against the estimated number of lives lost in order of decreasing LOL. Societal Risk Curves showing the limits of tolerability for existing dams, and new dams and major augmentations are shown in Figure 4-2 and Figure 4-3 below.

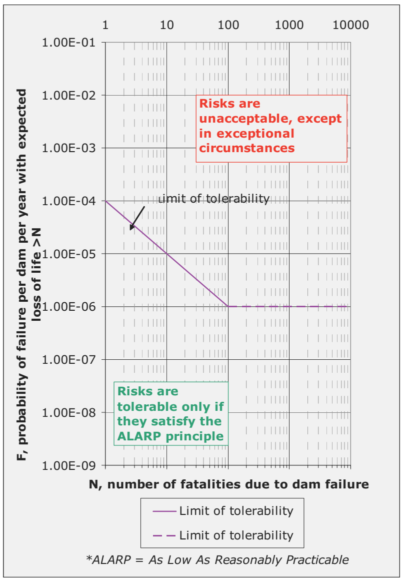


Figure 4-2 Revised ANCOLD Societal Risk Reference Guideline: New Dams and Major Augmentations

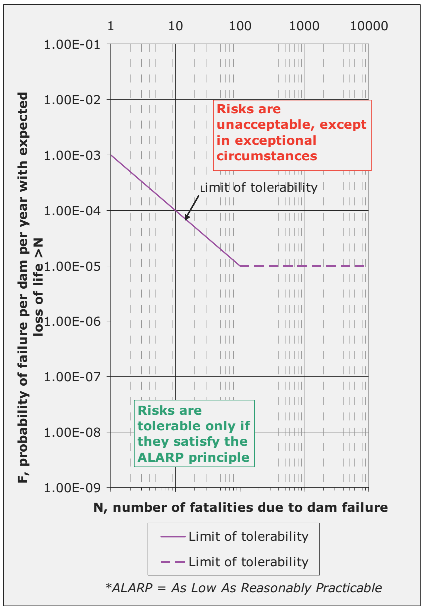


Figure 4-3 Revised ANCOLD Societal Risk Reference Guideline: Existing Dams

The tolerability of individual risk also varies depending on the nature of the structure being considered. For existing dams, an individual risk to the person or group, which is most at risk, that is higher than 10^{-4} per annum is unacceptable. For new dams or major augmentations of existing dams, an individual risk to the person or group, which is most at risk, that is higher than 10^{-5} per annum is unacceptable, except in exceptional circumstances.

Thus the risks contributed by all failure modes and scenarios need to be combined to obtain the overall risk to life in terms of both societal and individual risk. The chance of LOL per annum (individual risk) contributed by each earthquake, flood, and sunny day failure modes, is the product of:

- The annual probability of the event
- The conditional probability of dam failure, given the event
- The conditional probability of the individual's LOL, given dam failure (ANCOLD, 1998)

Using the ANCOLD Societal Risk Reference Guidelines, the 16 major and 14 minor dams included in PRA (2002) were ranked into risk categories for societal risk; being:

- Unacceptable risks which fall above the ANCOLD limit of tolerability and require high priority remedial works or investigations.
- Intermediate risks which fall under the ANCOLD limit of tolerability but above what was known as the ANCOLD Objective Line (in older revisions of the ANCOLD Societal Risk Reference Curve). These risks are subject to the principle of "As Low As Reasonably Practicable" (ALARP).
- Risks which are below both the ANCOLD limit of tolerability, and the ANCOLD objective line, and are tolerable only if they satisfy the ALARP principle.

The major dams, included in PRA (2002) were then separated into the above three categories and a group of SWC's 7 most critical dams, were formed. These dams are currently the subject of upgrading programs to satisfy the ANCOLD principle of ALARP. The dams included in the group of SWC's 7 most critical dams are Blowering, Burrendong, Chaffey, Copeton, Keepit, Split Rock, and Wyangala.

4.5.2 Economic Risk

In PRA (2002) economic risk was calculated as the sum of the products of the likelihood of dam failure in any one year, and the annualised economic damages for the full range of initiating events. In PRA (2002) the reporting of economic and financial risks followed the same principles as the reporting of risks to life, with "$" replacing "N", except that there is no parallel concept to that of the Maximum Individual Risk to Life. The tolerability of economic risks in SWC was not addressed in PRA (2002).

4.5.3 Environmental/Cultural Risk

In PRA (2002) environmental and cultural risk was represented by plotting the likelihood of failure in any one year, and the qualitative consequence ranking of pairs on environmental and cultural risk charts. Because the different categories of risks are incommensurable, risks cannot be aggregated across categories. However within a category, it may be appropriate to combine the raw "f" values, provided there is little variation in impact over the various failure scenarios. The tolerability of environmental or cultural risks in SWC was not addressed in PRA (2002).

4.5.4 Application of Consequences Assessments using Risk Database

As discussed above, consequence assessments including LOL, economic, environmental and cultural losses were assessed in PRA (2002) for the various cases of flood loading (Sunny Day, PMPDF with and without fail, DCF with and without fail) by consultants, SKM.

As estimates of the probabilities of failure for system components were derived for partitioned floods these ranges did not match those described in the consequences assessments. Interpolation of consequence assessments was required so that the results were appropriate for the level of flooding being considered, hence, results from the consequences assessments were not applied directly to the probability of failure numbers, but required considerable work to be interpolated.

Once each initiating event for every dam component had an associated appropriate set of consequences, this data was entered into a database, created internally within SWC, which could then be interrogated to derive risk data.

5 THE WAY FORWARD

5.1.1 General

As outlined above, PRA (2002) analysed 16 Major and 14 Minor dams within its portfolio and did not include Hume Dam or the Menindee Lakes Storage's. Since completion of this study, the minor dams have been transferred to the Department of Lands, and Fish River Water Supply has been merged with SWC hence Rydal and Oberon Dams need to be incorporated into the PRA. The PRA is now being progressed to include SWC's complete portfolio of dams prescribed under the Dam Safety Act which are listed in Table 2-1 above.

There have also been significant changes to the inflow hydrology of structures located in the area of the Generalised Tropical Storm Method Revised (GTSMR) application. Hydrology changes to inflow floods affect the areas of inundation determined by dambreak studies which in turn may alter the consequences, that may be associated with any failure event. Consequence assessments were previously undertaken in 2000 by consultants SKM. It is also possible that since this time developments have occurred within the inundated areas downstream of dams which may further impact 2000 consequences assessments. Overall, these changes may have a significant impact on the current risk level posed by dams in the region of the GTSMR. The changes required to review the current risk position of SWC are outlined below in terms of changes to both the probabilities and consequences of failure.

To move forward, SWC have also commenced a program of reviewing the revised risk position for each of the 7 most critical dams, post implementation of current proposed upgrading programs. This process will change the usefulness of the PRA for management and the Board in planning, particularly to prioritise proposed capital upgrading works. This phase of the PRA is outlined below and is known generally as developing the PRA.

5.2 Updating PRA (2002)

5.2.1 Revised Probabilities of Failure

Since completion of PRA (2002) the probabilities of failure have become outdated. To provide a current reflection of SWC current risk position, the following will be considered:

- Annual Exceedance Probabilities (AEP) of the inflow floods for dams in the GTSMR area, and the transition zone between the Generalised Southern Australia Method (GSAM) and the GTSMR areas have been revised, hence initiating events for dams in these areas should reflect these changes.
- Interim works have been completed at Keepit and Chaffey Dams. Changes to the structures relevant event trees and paths to failure should be revised accordingly.
- AEP's used to develop flood initiating events in PRA (2002) applied joint probability to the storage being at a level less than FSL, rather than the more conservative assumption accepted by the DSC that AEP's are derived from the storage initially at FSL. All dams to be included in the PRA should be reviewed to include initiating events commencing at FSL.
- In 2005 the probability of piping failure was revised using an event tree approach in lieu of the outdated UNSW Database method that was applied in 2002, this should be incorporated into the probabilities of failure for all dams in the prescribed portfolio.
- ANCOLD Guidelines on Risk Assessment, were revised in October 2003, hence these should be applied in revisions to the PRA.

5.2.2 Revised Consequences of Failure

Changes to hydrology in the GTSMR and the transition zone have a flow on effect to the consequences of failure associated with flooding events. Increased inflow floods increase the inundated areas downstream, hence possibly increase the population at risk and the LOL that may be expected.

PRA (2002) did not consider the effects of cascade failure at storage's with large dams upstream. Dambreak Studies and associated consequence assessments should consider these effects.

In PRA (2002), consequences for each of the partitioned initiating events were interpolated from the consequence assessments. This process was labour intensive, and included many assumptions. In moving forward with the PRA, a more practical approach is to run appropriate dambreak cases, and therefore evaluate consequences for individual flood partitions. This will provide a much greater level of accuracy, and be considerably more efficient.

Consequence assessments were undertaken for PRA (2002) by SKM and were largely completed in 2000. These are now upwards of 5 years old. Developments may have occurred downstream of dams, which may have an impact on the consequences associated with dam failure, even where there has been no substantial change to hydrology in the catchment area. Existing consequence assessments should be subjected to peer review particularly for currency.

The process to revise the consequences of failure has several stages, including the following:

- Update hydrology. Note that this task is discussed under probabilities above
- Revise the Dambreak Studies (including investigation of cascade effects where applicable)
- Review and compare existing consequence assessments. First, consideration should be given to ensure that previously developed consequences are still current (ie there have not been significant developments in the inundated area that were previously not identified). Secondly, a comparison should be made to determine if the existing consequence assessment data adequately covers the revised areas of inundation.

Changes in hydrology, as a result of changes to the GTSMR, therefore have a considerable impact on assessed consequences as a result of dam failure, and for this reason, SWC should reassess the consequences for dams in the GTSMR region as well as the transition zone. Dams in the GTSMR requiring consequence reassessment include Keepit, Chaffey, Pindari, Toonumbar Dam, Split Rock and Copeton Dams. Dams in the transition zone requiring consequence reassessment include Lostock, Glennies Creek and Glenbawn Dams.

The bulk of this work is required to comply with the statutory obligations of the DSC. To keep the PRA live and enable it to become a useful management tool, PRA (2002) should be revised and developed so that it is representative of SWC's current dam safety risk position, and hence useful as a management tool.

5.3 Developing the PRA (2002)

Together, the extended portfolio including revised probabilities and consequences will provide considerable data that SWC will be able to utilise to justify projects and programs to improve the dam safety risks associated with its large dam assets. This is one step to an improved risk understanding. Building upon this, SWC is currently developing PRA (2002) to analyse the risk position as a result of implementing upgrading works (staged where possible). This process will develop event trees for a staged program of work at each of SWC's 7 most critical dams. For dams where significant risk reduction may be achieved through undertaking upgrading works in stages, each stage of work, through to full deterministic compliance with DSC criteria, has been reviewed in terms of probabilities and consequences of failure.

The results of this analysis will provide the revised risk position (in terms of the ANCOLD Societal Risk curve) for the preferred option for each of the 7 most critical dams. This process is currently underway. It should be noted that each of these dams is currently at differing stages of progress toward upgrading, hence selection of a preferred option has been based on the available information at the time and may not reflect the final option selection. This process will allow SWC to compare the improvement to risk achieved at each dam with the implementation of staged/interim works, and ascertain the likely best value for money.

6 CONCLUSION

The goals of the current PRA exercise are two fold. SWC are currently reviewing the probabilities and consequences of failure, and in a parallel process, have also commenced a program to review the revised risk position as a result of implementation of upgrading works. Considerable resources will be required to deliver this extent of work, as well as work with management and the Board to ensure that the project is on track to deliver results that are useful, relevant and timely.

Many tasks to update the PRA, are justified to comply with the DSC, and hence require updating regardless of the existence of the PRA project. The challenge will be to ensure that studies required for statutory compliance, produce results with sufficient detail/information/data that may be used in the PRA.

SWC requires outputs of the PRA to enable it to understand and progress its business, and enable it to maintain and operate safe assets. As well as creating data that is useful to management, the PRA can be a useful tool to project managers of the 7 most critical dams to enable an understanding of issues of concern, and justify proposed works. It is important that adequate exposure and understanding is portrayed at all levels through SWC to enable further development of this tool in areas that will benefit the business as a whole.

7 REFERENCES

[1] ANG, A.S. and TANG, W. 1975, Probability Concepts in Engineering Planning and Design, John Wiley and Sons Inc. NY, 1975.

[2] Australian National Committee on Large Dams (ANCOLD) website

[3] Australian National Committee on Large Dams (ANCOLD) (1994) *Guidelines on Risk Assessment*, 1994.

[4] Australian National Committee on Large Dams (ANCOLD) (1998) *Guidelines on Risk Assessment, Position Paper on Revised Criteria for Acceptable Risk to Life*, August 1998.

[5] Australian National Committee on Large Dams (ANCOLD) (2003) *Guidelines on Risk Assessment*, October 2003.

[6] Dam Safety Committee (DSC) website

[7] Dam Safety Committee (DSC) *Information Sheet DSC2: Role, Policies and Procedures* (1999)

[8] FOSTER M.A., FELL R, DAVIDSON R, and WAN C.F. (2001*), Estimation of the Probability of Failure of Embankment Dams by Internal Erosion and Piping using Event Tree Methods*, ANCOLD Conference 2001.

[9] GRAHAM, W.J., (1999) *A procedure for Estimating Loss of Life Caused by Dam Failure DSO-99-06*, Bureau of Reclamation, September 1999.

[10] Seismicity Research Centre (SRC) (1999), *Review of Seismicity Southern/Northern NSW*, March 1999.

[11] SKM (2001), *24 Dams Portfolio Risk Assessment (Consequences Only)* 2000/2001.

[12] STANDARDS AUSTRALIA (1999), AS/NZS 4360:1999 Risk Management.

[13] SWC (1999) *Status of Risk Assessment of State Water Dams: Screening Level Risk Assessment DLWC*, State Water, August 1999

Acknowledgments

The authors gratefully acknowledge the assistance of colleagues in the Department of Commerce, and the support and cooperation of State Water Corporation in allowing us to present this paper.

HELICOPTER GEARBOX BEARING BLIND FAULT IDENTIFICATION USING A RANGE OF ANALYSIS TECHNIQUES

Sawalhi N* and Randall R.B*

* *School of Mechanical and Manufacturing Engineering, The University of New South Wales, Sydney 2052, Australia*

Abstract: Vibration acceleration signals were obtained from an overload test of a Bell 206 Helicopter Main Rotor Gearbox in order to complete a blind bearing fault analysis where no knowledge of the fault was made available prior to the analysis. A range of diagnostic techniques was applied. These included constant percentage bandwidth (CPB) spectrum analysis, spectral kurtosis (SK) analysis to determine the frequency bands with maximum impulsivity and to filter the signal to maximize that impulsiveness, and envelope analysis to determine the fault frequencies. Order tracking was used to compensate for speed fluctuations, while linear prediction using autoregressive models (AR) was used to remove the regular gear meshing contribution in the signals. As a result of applying these techniques, a fault in one of the planetary bearings was identified. A match with the cage frequency and the inner race ball pass frequency indicated deterioration associated with these components. Roller fault frequencies were not directly detected, but the fact that roller faults give a modulation at cage frequency shows that their effect was still detected. SK gave a good measure of the severity of the fault when compared to the amount of metal wear debris in the oil. Details of the test, as well as application of a statistical fault detection technique can be found in a companion paper submitted by the Defence Science and Technology Organization (DSTO) Australia.

Key Words: Order tracking, Autoregressive models, Spectral kurtosis, Complex Morlet wavelets, Constant percentage bandwidth.

1 INTRODUCTION

Many failures which occur in rotating components like gears and bearings often show their signature in the vibration signal, and can be detected in the early stages. Monitoring such vibrations in helicopter gearboxes often requires an extensive interpretation by a trained diagnostician. This is due mainly to the existence of a large number of rotating components, all of which contribute and mix in different ways [1], making it very hard to track changes in a certain component.

Helicopter transmission gearboxes have been investigated in detail to understand the types of baseline frequencies and failure indicators that can be detected by monitoring vibrations [2,3]. These efforts have resulted in a better understanding of the way in which different signals express themselves and mix with others, and have led to the development of a variety of diagnostic and prognostic techniques [4,5,6].

Different techniques have been utilized for bearing diagnostics in such environments, aiming at separating the gear signals from the bearing ones. These include the use of synchronous averaging [4], Self-Adaptive Noise Cancellation (SANC) [7], and Discrete Random Separation (DRS) [8]. Once the separation is achieved, envelope analysis [9] - which is a widely used technique for bearing analysis - is employed to extract bearing defect frequencies, thus indicating the source and the nature of the fault. However, as minor speed variation exists in practice, it is necessary to remove this variation and phase-lock the vibration signal to shaft speed. This is achieved through the use of order tracking, and is usually performed by re-sampling the vibration data at uniform rotational angles - using a tacho signal - rather than time based intervals [1]. In reference [10] Bonnardot et al showed that this re-sampling can be achieved using the signal itself, and combined that with the use of self adaptive noise cancellation, which resulted in a clear identification of a ball fault in a planetary bearing and proved to be much more efficient than using SANC on its own. However, the selection of the frequency band for demodulation purposes (envelope analysis) was performed on the basis of trial and error, and needed to test more than one band to be able to detect the fault.

In this paper, different techniques are combined to provide an effective and automated extraction of a fault from a planetary bearing, which was blindly analysed from accelerometer signals measured at the helicopter transmission test rig facility at DSTO. The proposed algorithm starts by compensating for speed variations by re-sampling the signal synchronously with rotational speed using a tachometer signal. The re-sampled signal is then analysed using a newly developed technique for rolling element bearings [11]. This technique combines the use of autoregressive models, complex Morlet wavelet banks and spectral kurtosis (SK) [12,13] to automate the selection of the best band for envelope analysis.

The algorithm is taken a step further, by investigating its suitability for trend analysis, and possibly the prognosis of the fault. This is made possible by comparing the outcome of the algorithm (Spectral kurtosis evolution and filtered signals) with the amount of metal debris obtained from the test rig lubricant, and also by comparing constant percentage bandwidth (CPB) spectra at different stages of the fault.

2 EXPERIMENTAL

Experimental data was obtained from a test performed in the Defence Science and Technology Organization (DSTO), Bell 206 test rig facility in Melbourne, and was blindly analyzed for a defect in one of the bearings in the transmission system. The conditions for generation of the data included an overload test, which was carried on over different time intervals allowing the monitoring of fault development at different stages.

2.1 Bell 206 Transmission Description

The Bell 206-transmission system as shown in Figure 1 is a two-stage reduction gearbox. The first stage consists of a spiral bevel pinion gear with ($N_{pinion} = 19$ teeth) driven by the input shaft from the engine side and rotating at a speed of ($f_{input} = 100$ Hz), which meshes with a bevel gear ($N_{bevel} = 71$ teeth). Triplex ball bearings and one roller bearing support the bevel pinion shaft. Duplex ball bearings and one roller bearing support the bevel gear shaft in an overhung configuration.

The second reduction stage is provided through a planetary mesh. The epicyclic gear system consists of a sun gear ($N_{sun} = 27$ teeth), splined to the bevel gear shaft, which in turn drives three planet gears ($N_{planet} = 35$ teeth). The planet gears mesh with a ring gear ($N_{ring} = 99$ teeth), which is attached to the top casing. Power is transmitted through the planet carrier, which is attached to the mast output shaft. The overall reduction ratio of the main transmission is (17.79:1) driving the output at ($f_{out} = 5.73$ Hz).

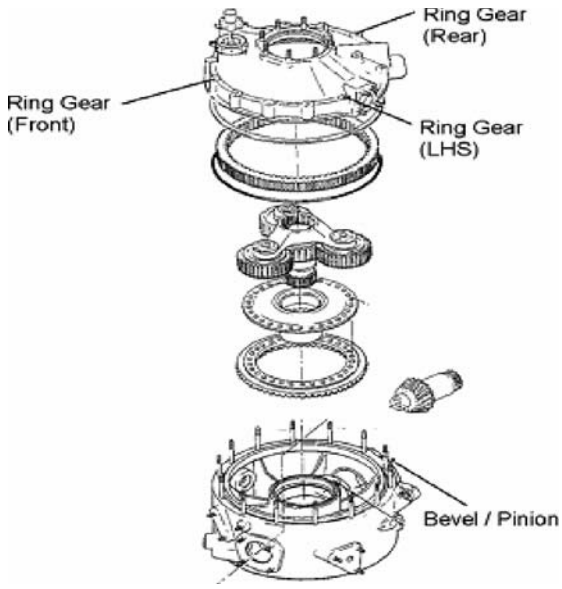


Figure 1 Bell 206 Transmission

Vibration data was collected from four accelerometers located on the transmission housing as shown in Figure 1. Three of them were attached to the ring gear, measuring vibration at the second stage (planetary gears). At a sampling frequency of 51.2 kHz, 30-second records were obtained at different stages of the fault. Each accelerometer is mounted so that its axis of measurement intersects the rotation axis of the shaft at 90 degrees. In addition to the accelerometers, two tachometers were placed at the input and output shafts. For this analysis, only one accelerometer was used as the one that gave the best indication of the fault. This was the front accelerometer.

Debris was collected from the gearbox oil using an inline MetalSCAN unit. This was done in parallel with vibration measurements, and the cumulative mass was recorded against the test hour. Oil temperature was also monitored all the way through the whole test.

2.2 Expected Frequencies

The vibration signal is expected to contain all frequencies due to meshing between the different sets of gears, their harmonics and sidebands. References [5,14] give a complete description of computing those frequencies. This is adopted here for the same purpose, taking into consideration the differences in rotational speeds and the numbers of teeth on the different gears. Figure 2 shows a schematic presentation of the planetary gear system. The motion of such a system can be easily understood by a superposition analysis, which can be done in two sequential steps and summed to obtain the relative movements between the different components [14].

In the first step, the carrier (arm) is “locked” in position and the ring gear is free to rotate. The planet gear and the sun gear rotations are estimated based on that.

In the second step all elements of the gear train are "locked" together and the entire assembly is rotated backwards until the ring gear is back to its original position.

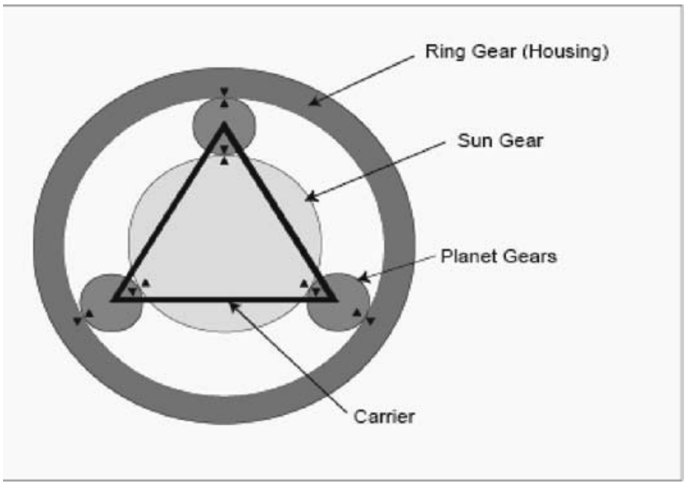


Figure 2 Elements of a simple planetary or epicyclic gear train [14]

Summing those two actions provides the total rotations of all components. The relevant values for the system under investigation are shown in table 1.

Table 1

Shaft speed estimation for different rotating parts

Rotational frequency* (Hz) of: -	Planet Gears f_{planet}	Sun gear f_{sun}	Gear carrier (Output shaft) f_{out}	Planet gear with respect to planet carrier (spin frequency) f_{spin}
	-10.48	26.76	5.73	-16.21

* Values estimated based on 100 Hz input shaft frequency.

Note that, as the planet gears and the carrier rotate in opposite directions, the net speed of a planet gear with respect to the carrier is the sum of the absolute speeds of the planet gear and the carrier. This is known as the spin frequency of the planetary gears, and can be alternatively calculated using equation (1).

$$f_{spin} = -\frac{N_{ring}}{N_{planet}} f_{out} \qquad (1)$$

where the negative sign in equation (1) indicates that the planets spin in the opposite direction to the sun gear and the carrier. Results showing shaft speeds and gear mesh frequencies in the Bell 206 transmission system are presented in table 2.

Table 2

Gear, mesh frequencies		
	Frequency of interest	Hz
Stage 1	Input shaft frequency	100.00
	Pinion Mesh frequency	1900.00
	Bevel Gear shaft speed	26.76
	Bevel Mesh frequency	1900.00
Stage 2	Sun mesh frequency	722.54
	Carrier frequency (output arm speed)	5.73
	Epicyclic mesh frequency	567.71
	Planet pass frequency	17.20

Rolling element bearing frequencies are generated when a bearing begins to fatigue. These will be too low to show up in the early stages of the fault due to the strong contribution from the meshing frequencies. For a fixed outer race (most common situation), the load direction is fixed with respect to the outer race, and equations (2 – 5), [15] are used to calculate the defect frequencies. Note that this is kinematically identical to the planet gears discussed above for a load angle $\alpha = 0$. The load angle is taken for the dominant load path, but actually varies for each rolling element, which is the reason for the random slip that always occurs. Thus, actual mean bearing frequencies typically differ from the theoretical ones by 1-2%, and the corresponding pulse period also varies randomly by about the same amount.

Inner Race Defect Frequency [BPFI]

$$f_{BPFI} = \frac{Nf_i\left(1 + \frac{d}{D}\cos(\alpha)\right)}{2} \tag{2}$$

Outer Race Defect Frequency [BPFO]

$$f_{BPFO} = \frac{Nf_i\left(1 - \frac{d}{D}\cos(\alpha)\right)}{2} \tag{3}$$

Cage Rotational Frequency relative to outer race (fundamental train frequency) [FTF]

$$f_{FTF} = \frac{f_i\left(1 - \frac{d}{D}\cos(\alpha)\right)}{2} \tag{4}$$

Ball/Roller rotational speed around its axis (ball/roller spin frequency) [BSF]

$$f_{BSF} = \frac{f_i}{2}\frac{D}{d}\left(1 - \left(\frac{d}{D}\cos(\alpha)\right)^2\right) \tag{5}$$

where N is the number of rolling elements, d and D are the rolling element diameter and the pitch diameter of the bearing respectively. f_i is the rotational speed of the inner race (shaft speed). α is the contact angle (from the radial).

Planetary bearings represent a special case, as the bearing is attached to the planet gear on its outer race, and to the carrier on its inner race, which means that both races will be rotating. Equations 6-10 are used to calculate the defect frequencies for the planetary bearings. Note that the load direction in the planetary bearings is fixed with respect to the inner race rather than the outer race (FTF is calculated as the cage speed relative to the inner race (carrier)), so that it is BPFO that is modulated by passage through the load zone, rather than BPFI as in the normal case.

Inner Race Defect Frequency [BPFI]	$$f_{BPFI} = \frac{N(f_{planet} - f_{out})\left(1 + \frac{d}{D}\cos(\alpha)\right)}{2}$$	(6)
Outer Race Defect Frequency [BPFO]	$$f_{BPFO} = \frac{N(f_{planet} - f_{out})\left(1 - \frac{d}{D}\cos(\alpha)\right)}{2}$$	(7)
Cage Rotational Frequency (absolute)	$$f_{cage} = \frac{f_{planet}\left(1 - \frac{d}{D}\cos(\alpha)\right)}{2} + \frac{f_{out}\left(1 + \frac{d}{D}\cos(\alpha)\right)}{2}$$	(8)
Cage Rotational Frequency relative to inner race (carrier) [FTF]	$$f_{FTF} = \frac{(f_{planet} - f_{out})\left(1 + \frac{d}{D}\cos(\alpha)\right)}{2}$$	(9)
Ball/Roller rotational speed relative to both races (ball/roller spin frequency) [BSF]	$$f_{BSF} = \frac{f_{planet} - f_{out}}{2}\frac{D}{d}\left(1 - \left(\frac{d}{D}\cos(\alpha)\right)^2\right)$$	(10)

where N is the number of rolling elements, d and D are the rolling element diameter and the pitch diameter of the bearing respectively. f_{planet} and f_{out} are the rotational speeds of the planetary gears and the output shaft (carrier) as in table 1. α is the contact angle (from the radial).

Table 3 presents the theoretically calculated defect frequencies for bearings in the Bell 206 transmission system.

Table 3

Bearing defect frequencies	Shaft speed (Hz)	(BPFI) (Hz)	(BPFO) (Hz)	(FTF) (Hz)	(BSF) (Hz)
Triplex bearing - Input pinion	100.00	822.72	577.28	41.23	276.43
Roller bearing - input pinion	100.00	616.58	383.42	38.34	202.79
Bearing lower mast	26.76	273.63	208.85	11.56	96.45
Duplex bearing, input gear shaft - lower support	26.76	369.22	299.80	11.99	127.55
Duplex bearing, input gear shaft - upper support	26.76	356.68	285.57	11.90	59.85
Planetary bearings		117.72	76.95	9.81	37.04

3 DATA PROCESSING AND FAULT DIAGNOSTICS

The data from the accelerometers was processed in a number of sequential steps as shown in the block diagram of Figure 3. The logic and ideas behind each processing step are as follows: -

1. Minimizing the speed fluctuations that may exist in the transmission system.

   As has been indicated in the Introduction, the system under consideration contains a large number of rotating components, which all contribute to the vibration signal. Although the system in general runs at a nominal constant speed, it will exhibit minor speed variation in practice. For this reason, sampling of the vibration data is performed at rotational rather than time-based intervals [1]. This angular resampling (also known as order tracking) will force the gear signal to be deterministic and is done here using the tacho signal.

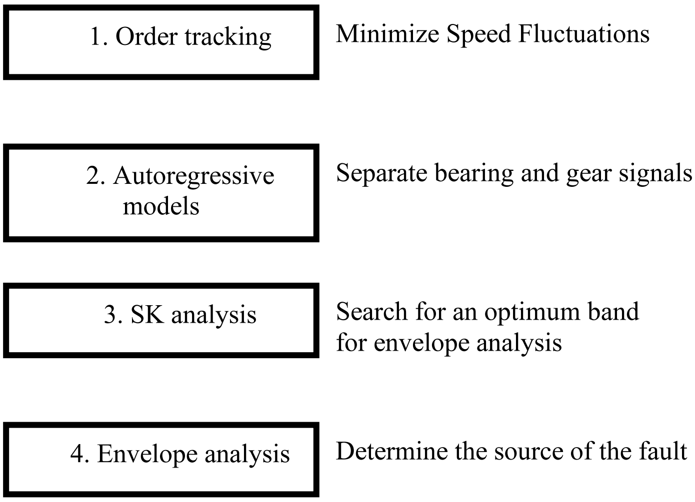


Figure 3 Block diagram presentation of signal processing steps

2. Separating bearing signals from gear signals and pre-whitening.

   This is done using autoregressive models (AR), with the aim of analyzing the residual signal of the model (the so-called pre-whitened signal). The use of this signal in further processing steps, is more favourable than the use of the original signal (order tracked), and proved to give better results for the spectral kurtosis analysis [16].

3. Searching for an optimum band for demodulation purposes.

   This is done using complex Morlet wavelets. These are used as a filter bank with constant proportional bandwidth (uniform resolution on a logarithmic frequency scale). Different banks are used to select the best filter for the envelope analysis as the one that maximizes the SK

4. Extracting bearing defect frequencies and determining the source of the fault.

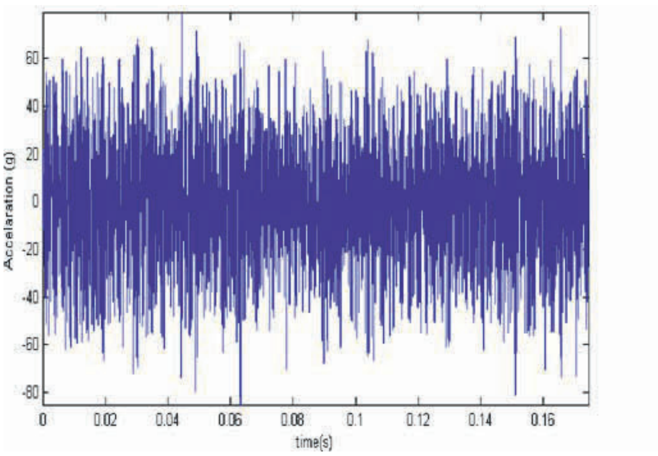


Figure 4 Accelerometer signal representing 1 rotation of the carrier

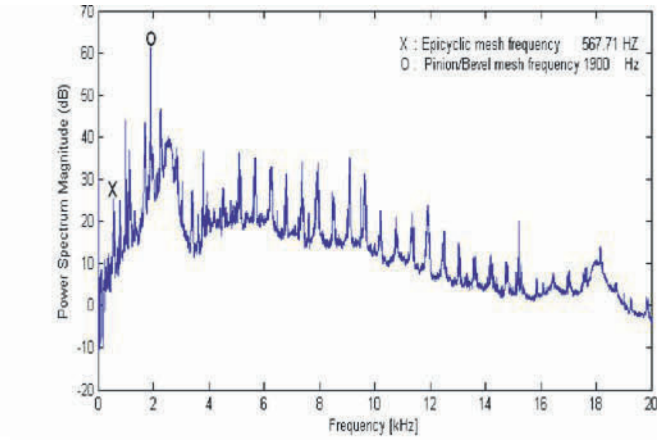


Figure 5 Power spectral density of the raw signal

For the sake of diagnostics, only the result from the last measurement is discussed here, while more discussion is given on the fault in its different stages in the prognostic section. A depiction of the first 8,930 samples (174.36 ms) representing the first rotation of the carrier (5.73 Hz) is given in Figure 4. With a kurtosis value of (-0.61) it is very difficult to recognize any impacts in the time domain signal. This is further investigated by examining the power spectral density (PSD) of the signal as shown in figure 5, which shows how the harmonics of the gear mesh frequencies (two families) dominate the spectrum, and can be seen over the whole useful range of the signal (to 20 kHz).

3.1 Order tracking

The main idea behind the order tracking is the removal of speed fluctuations that can exist in the transmission system leading to a non-deterministic gear signal. Order tracking is a way of ensuring the deterministic property of the gearmesh signal while facilitating its treatment and interpretation in further steps of the analysis. Order tracking in concept is based on re-sampling the signal in the angular domain rather than the time domain. This can be done either by using hardware (phase-locked frequency multiplier), or by digitally re-sampling the signal with the aid of a tachometer signal [1]. Recently this was done by using the signal itself, based on phase demodulation of the gearmesh frequencies [10].

In this study, all order tracking was performed using a digital re-sampling technique, where both the vibration signal and the tachometer signal were sampled simultaneously using a fixed time-based sampling frequency. Matlab® code was used for this purpose and included the use of a cubic spline interpolation of the vibration signal to calculate the values at the required sample points, based on the tachometer signal.

3.2 Linear Prediction filtering and pre-whitening based on Autoregressive modelling

Having obtained an order-tracked signal, with a deterministic gearmesh contribution, autoregressive models (AR) are used next with two main aims. The first is to remove gear contributions from the signal. This will result in the separation of the signal into two main parts, namely, a deterministic part (AR fitted signal – containing the gearmesh contribution) and a noise part (containing additive white noise and nonstationarities (impulses)). Because of the randomness referred to above, the bearing signals are contained in this part [7].

The second aim from using AR models is pre-whitening. The noise part of the AR model is close to that of a white noise (power spectral density changes with frequency are very small), as it is made up of impulses and additive white noise. This has the advantage of solving an anomaly in the SK, and gives better results when used prior to the application of the SK [16].

In the process of designing an AR filter for linear prediction, a value of the signal at time index k is normally based upon a linear combination of prior values, according to equation (11) [17]:

$$x_k = -\sum_{i=1}^{p} a_i x_{k-i} + \varepsilon_k \tag{11}$$

where the a_i are the autoregressive coefficients and ε_k is the residual error. The residual error ε_k (difference between the actual signal and the AR prediction) has an almost white spectrum, but can be made up of stationary noise and impulses.

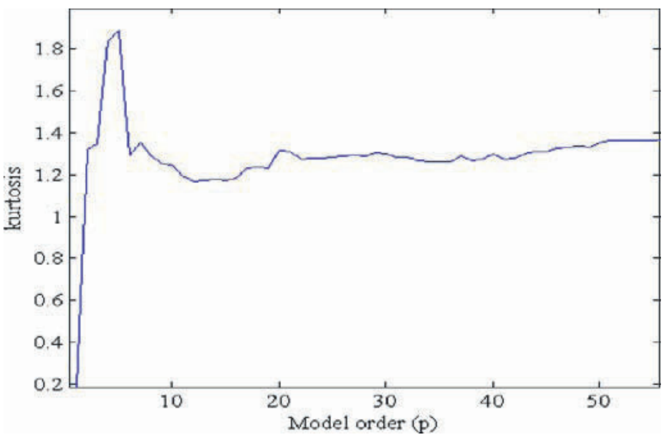


Figure 6 kurtosis of residual signals for different model orders

The AR filter as described by equation (6), well predicts the deterministic pattern of the signal but is not capable of adapting to the sudden impulses caused by a localized fault. As a consequence, the fault impulses will be left in the residual

ε_k of the AR filter. If SK is applied on the residual signal, high SK values and a clearer extraction of fault impulses are obtained.

Several criteria are in place to select the optimum model order, such as the Akaike Information Criterion (AIC) [18], however as the main objective of using the AR technique for rolling element bearings resides in separating the impulses from the original signal, it is intuitive to base the order selection criterion on maximizing the kurtosis of the residual signal, which as has been indicated earlier, contains both the impulses and the stationary noise. The optimum model order p should be less than the spacing between two consecutive impulses, as this guarantees that the model does not adapt to the impacts as being a part of the deterministic signal and that they will be contained in the residual signal.

The optimum order for the AR model was first examined based on maximizing the kurtosis of the residual signal. The result as depicted in figure 6 clearly shows that a filter length of five samples would be optimum and is preferred for a better extraction of impacts hidden in the system.

Figure 7 shows the effect of using a linear prediction filter and the benefits of analysing the residual signal rather than the raw signal. Impulses throughout the signal can be seen very clearly in the residual signal (b) when compared to signal (a) from the same figure. The passage of the planet gears can also be seen clearly (three bursts of increased amplitude in one rotation) when inspecting one rotation of the carrier as presented in figure 8. Other impulses do exist in the signal and will require further analysis to determine the source. A final point to mention in this regard is the increase of the kurtosis value from – 0.61 (in the order-tracked signal) to about 2.2 in the residual signal.

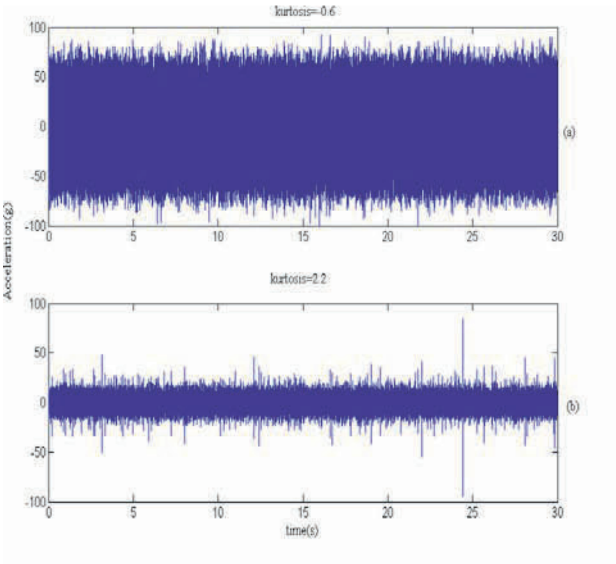


Figure 7 (a) Initially processed signal by order tracking (b) AR (5) residual signal

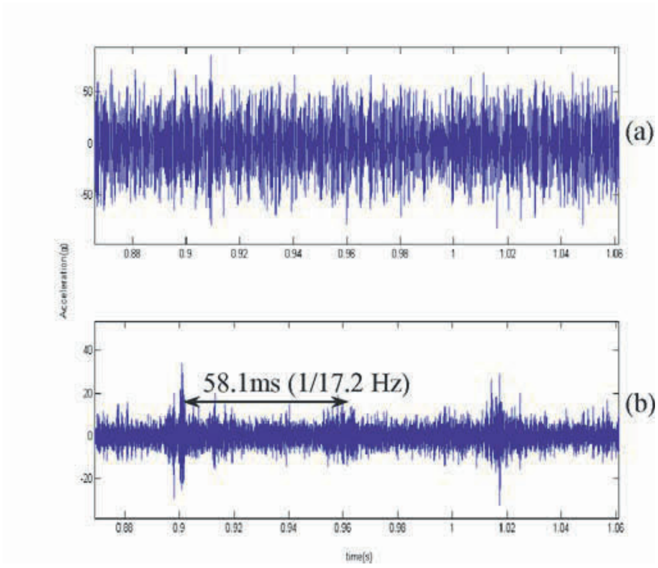


Figure 8 Zoom-in views (one rotation of the carrier) showing the effect of using an AR linear prediction filter for signal separation, (a) Order tracked signal, (b) Residual signal

3.3 Automated envelope analysis using Spectral Kurtosis and complex Morlet wavelet filter banks

The presentation of this section is based on what has been presented in [11], in which Sawalhi and Randall used different filter banks constructed using Complex Morlet wavelets to select the best filter for the envelope analysis - in terms of centre frequency and bandwidth - as the one that maximizes the SK. This approach proved to be well suited for the purpose of rolling element signal analysis due to:

- The use of wavelets being more favourable than the STFT (short time Fourier transform) as the duration of the wavelet varies inversely with frequency corresponding to constant relative damping.

- The choice of complex Morlet wavelets helps to design a filter, which acts as a ‘matched filter’ to condition transient signals in the data.
- The complex wavelets help detect transients in a uniform manner regardless of their phase.

The complex Morlet wavelet is defined in the time domain as a complex exponential wave multiplied by a Gaussian function, and has the shape of a Gaussian function in the frequency domain as follows:

$$\psi(t) = \frac{\sigma}{\sqrt{\pi}} e^{\sigma^2 t^2} e^{i2\pi f_o t} \tag{12}$$

$$\Psi(f) = \Psi^*(f) = e^{-(\pi^2/\sigma^2)(f-f_0)^2} \tag{13}$$

where Ψ(f) is the Fourier Transform of ψ(t). Since Ψ(f) is real, Ψ(f)= Ψ*(f), where * denotes the complex conjugate. f_0 is the centre frequency of the window and σ determines its width. Since the real and imaginary parts of ψ(t) are Hilbert transforms, it is analytic and its Fourier transform Ψ(f) is one-sided with positive frequencies only.

The advantage of the complex Morlet wavelets, compared with the real version using windowed cosines, is that the imaginary parts (sines) have their maximum when the cosines have zero crossings, so that the squared amplitude of the wavelet coefficients is not sensitive to the phasing of local features in the time signals.

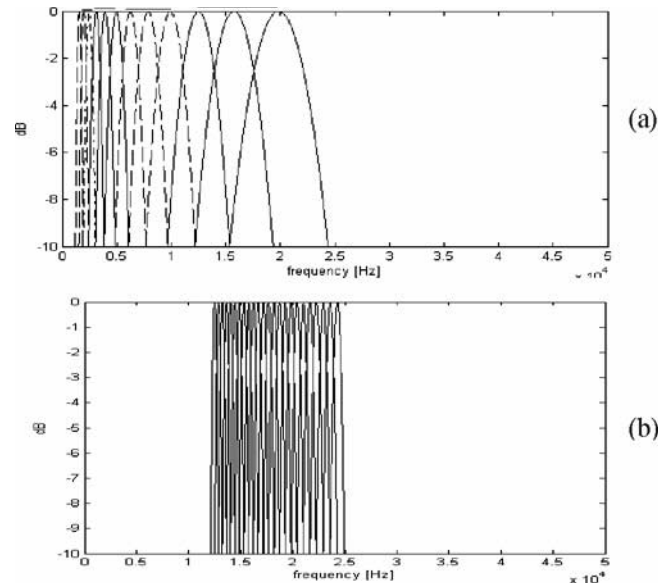


Figure 9 Morlet wavelet filter banks used in [12] (a) 1/3-octave over four octaves (b) 1/24 octave over an octave

To optimise the use of complex Morlet wavelets as a filter bank, and to gain a balanced representation of the signal in time and frequency domains, a few banks with different numbers of filters per octave e.g. (3,6,12,24) or (2,4,8,16) were suggested to decompose the signal. In order to gain efficiency, the number of octaves covered in each bank can be reduced as the number of filters per octave is increased, as the coarser analysis gives an indication of the frequency bands with the highest impulsiveness. Moreover, filters for lower octaves can be derived by progressively down-sampling the higher octave filters by factors of 2, which accelerates the analysis. Figure 9 shows two filter banks constructed using complex Morlet wavelets. A coarse one (3 filters/octave) spreading over 4 octaves and a fine one (24 filters/octave) over one octave.

Once the filters are constructed, the signal can then be decomposed correspondingly. This is easily performed by multiplying the FFT of the signal by each row of the filter bank matrix, then by performing the IFFT to obtain the wavelet coefficients (complex values) [11]. The multiplication of the wavelet coefficients by their complex conjugate in fact gives the squared envelope of the filtered signal. If y is the squared envelope, then the kurtosis corresponding to y can be calculated using equation (14), to give a single value for each filter.

$$kurtosis(y) = \frac{mean(y^2)}{(mean(y))^2} - 2 \tag{14}$$

The value 2 is subtracted here to obtain a value of zero for the squared envelope of Gaussian noise. The best filter is chosen as the one that gives the highest value of kurtosis, i.e. maximizing the SK from the squared envelope. This filter is retained and then re-used to obtain the squared envelope signal and the envelope spectrum.

After processing the residual signal (figure 6(b)) using the set of filter banks described, the result is plotted in figure 10 as an image, whose rows represent the filter banks (3,6,12,24 filters/octave), and columns represent the filters in each bank. The values assigned to each slot (filter) are the kurtosis of the squared envelope signal for that specific filter in the bank, and were calculated as shown in equation (9). The highest kurtosis indicates the optimum slot (filter) for the purpose of envelope analysis. The maximum SK was obtained using 12 filters/octave (centre frequency of 18800 Hz and bandwidth 1175 Hz).

To gain more insight into the excitation across the frequency range, the SK is plotted in figure 11 using a bank of 12

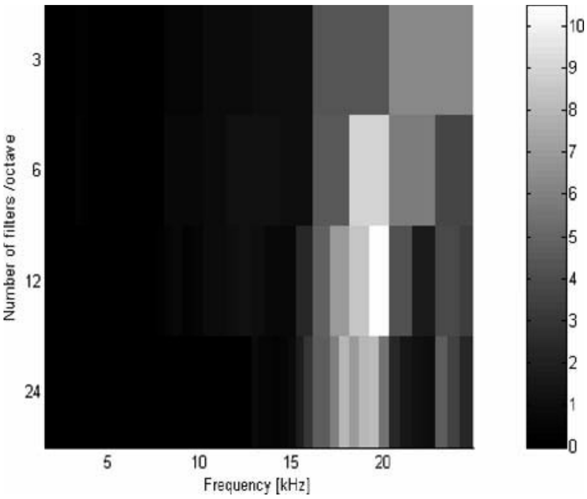


Figure 10 Image representation of SK variation for 4 filter banks; namely (3,6,12,24) filters/octave

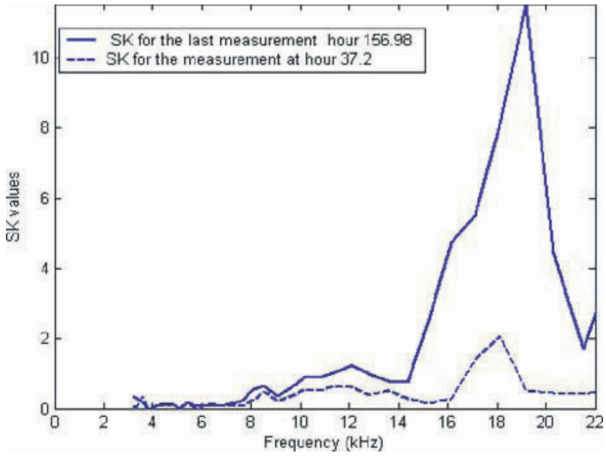


Figure 11. SK variations between first measurements at hour 37.2(dotted) and last measurement at hour 157

In figure 13, the harmonics of the carrier are highlighted. It is quite clear that the planet pass frequency (3 times carrier speed) and its harmonics show strongly in the spectrum. These are indicated by (x) in the same figure.

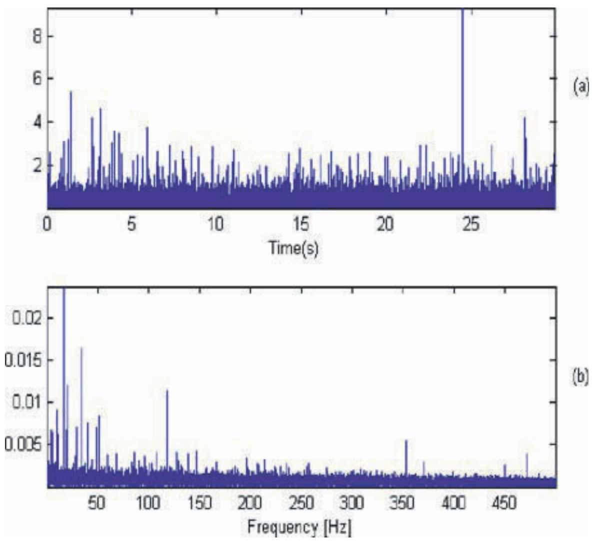


Figure12 (a) Squared envelope and (b) envelope spectrum of the filtered signal *(last measured signal)*

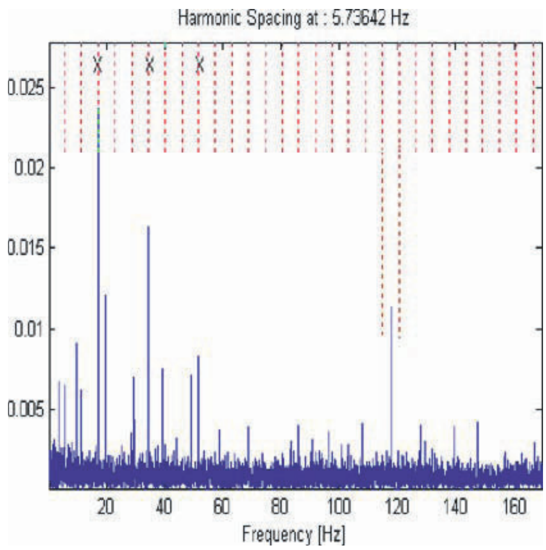


Figure13 Envelope spectrum showing the harmonics of the carrier - (dotted)- and the planet pass frequency as indicated by (x)

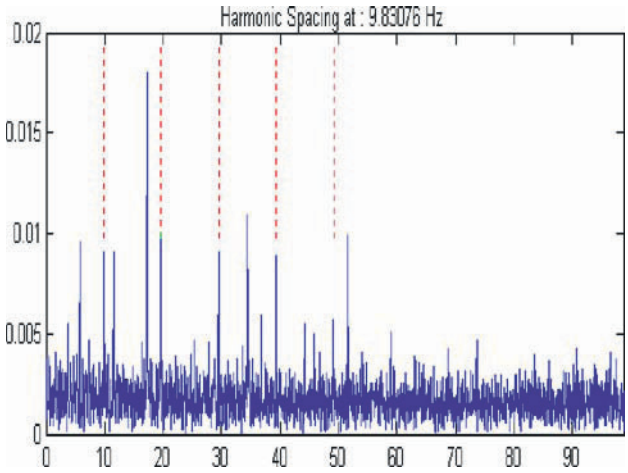


Figure14 Squared envelope spectrum showing the harmonics of 9.8 Hz

Figures 14 and 15 show the other dominant frequencies that exist in the envelope spectrum. A family of frequency spacing 9.83 Hz, shown in figure 14, and a family of spacing 117.69 Hz, shown in figure 15.

In Section 2.2 of this paper, interesting frequencies and bearing defect frequencies were calculated. These have been consulted at this stage to see if a match exists with the frequencies appearing in the envelope spectrum.

A search for 9.83 Hz and 117.69 Hz was conducted and a match was found with the frequencies of one of the bearings in table 3. This is the planetary bearing, which is a two-row bearing. The frequency at 9.83 Hz matches its cage frequency relative to the carrier (FTF), while the 117.69 matches that of the inner race (BPFI).

Figure 16 shows the sidebands around 117.69, which are related to the 9.83 Hz. This is because of the interaction between the inner race fault and the roller faults

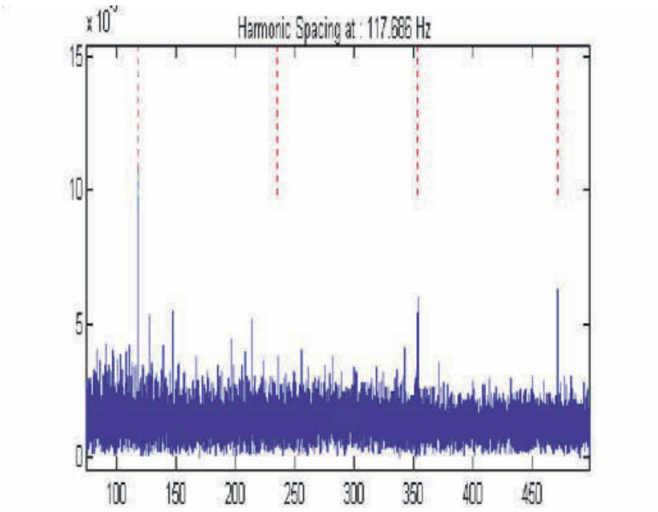


Figure15 Squared envelope spectrum showing the harmonics of 117.69 Hz

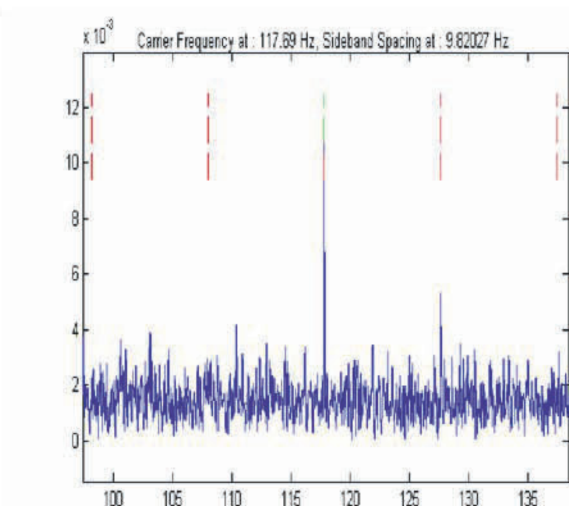


Figure 16 Sidebands around the 117.69 Hz

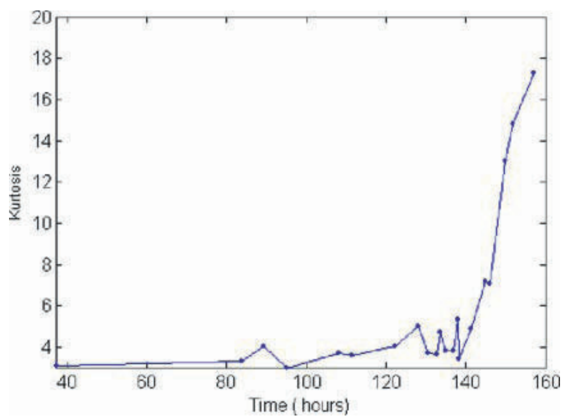


Figure 17 Kurtosis of filtered signal based on the optimum filter stated by the SK analysis

This is a clear indication of the source of the excitation sensed by the SK. *A defective planetary bearing* exists in the system. The existence of the (FTF) and its harmonics in the envelope spectrum, as seen in figure 14, and the BPFI with harmonics and sidebands – figures 15 and 16 - indicates a variation at cage frequency (usually due to differences between the rolling elements) and localised inner race faults.

4 PROGNOSTICS AND FAULT SEVERITY INDICATORS

One of the key issues in bearing prognostics is the ability to detect the defect in its incipient stages, which in turn gives enough time to track its development and avoid any catastrophic failure. SK armed with recent optimisations and enhancements gives a clear separation of the signal arising from a bearing defect, and thus is ideal for being implemented in prognostic models. It gave a good measure of the severity of the fault (figure 17) when compared to the amount of metal wear debris in the oil (figure 18).

Constant percentage bandwidth (CPB) comparisons were also performed to see how the fault severity is tracked using this technique. In the CPB analysis, as the name suggests, spectra used for comparisons are of constant percentage bandwidth (CPB), rather than of constant bandwidth and cover a wide frequency range of three decades.

Figure 19 shows a 3D plot of the CPB comparisons (linear frequency scale), showing where the most excited bands are, and how this evolves with time i.e. as the bearing condition gets worse. A clear increase of the dB difference is seen around 18 kHz frequency (which matches the results earlier presented through the analysis of SK). However, neither the values, nor the pattern indicates the severity sensed by the accumulated mass of wear detected in the oil, or the increase in the envelope spectrum as presented in the diagnostic section between the start of the test and the final measurement. A significant change is normally 6 dB or more. The CPB spectrum comparison gave a weak indication of the deterioration but could not be considered significant. There was also little change in the temperature of the oil (figure 20).

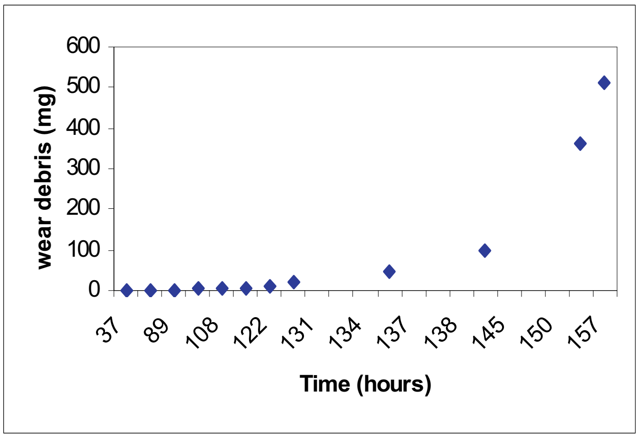


Figure 18 Accumulated wear debris as test proceeded

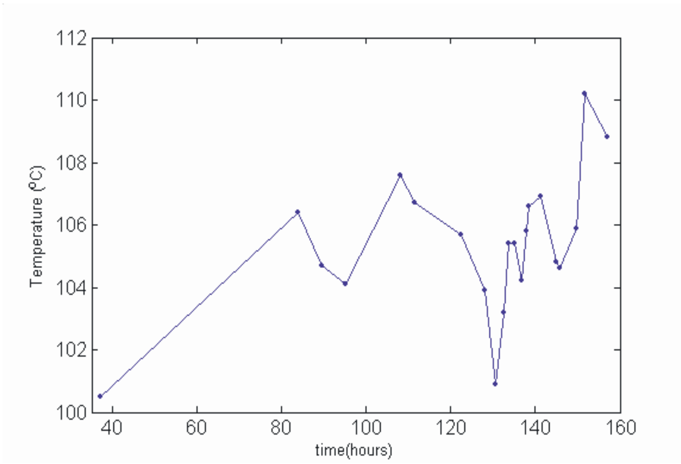


Figure 20 Oil Temperature variations as test proceeded

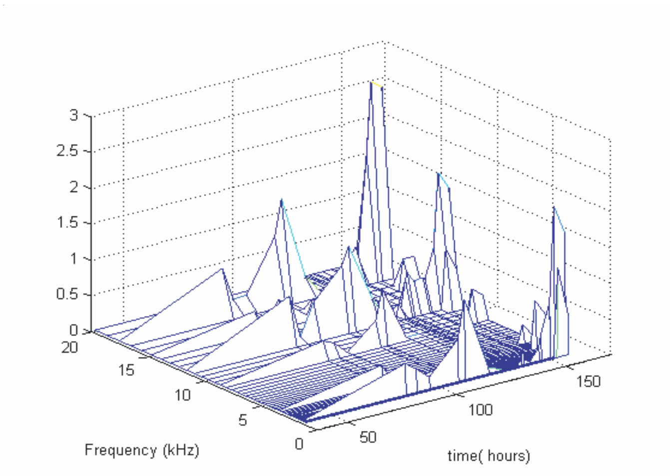


Figure 19 Constant Percentage Bandwidth (dB) - Reference measurements at hour 37.2

Figure 19 shows a 3D plot of the CPB comparisons (linear frequency scale), showing where the most excited bands are, and how this evolves with time i.e. as the bearing condition gets worse. A clear increase of the dB difference is seen around 18 kHz frequency (which matches the results earlier presented through the analysis of SK). However, neither the values, nor the pattern indicates the severity sensed by the accumulated mass of wear detected in the oil, or the increase in the envelope spectrum as presented in the diagnostic section between the start of the test and the final measurement. A significant change is normally 6 dB or more. The CPB spectrum comparison gave a weak indication of the deterioration but could not be considered significant. There was also little change in the temperature of the oil (figure 20).

5 FAULT INSPECTION AND VERIFYING THE FINDINGS

A good result was achieved for the analysis of the test data. The fault was in one of the planet gear support bearings as has been indicated in the analysis (figure 21.a). Furthermore there was damage to the inner race (figure 21.b) correctly identified, and to the rolling elements as well (figure 21.c). The fault, which developed in the inner race and the rolling elements, showed its signature clearly in the envelope spectrum through the harmonics of the inner race pass frequency, which were modulated by the load variation frequency in the last stages, indicating severe damage. SK gave an excellent measure of the severity of the fault and the bearing degradation process was clearly represented by its evolution. Other tested parameters for severity analysis such as CPB analysis and vibration statistics of the raw signal gave little indication of the fault. There was also little change in the temperature of the oil. The end result of the processing algorithm (envelope spectrum) contains all the necessary information and the whole algorithm can be automated for a complete identification of any bearing related faults. This is in contrast to the traditional diagnostic techniques, which require the analyst to select a suitable band for performing the envelope analysis (requires historical data or can be done by trial and error), and may need more than one position to measure at.

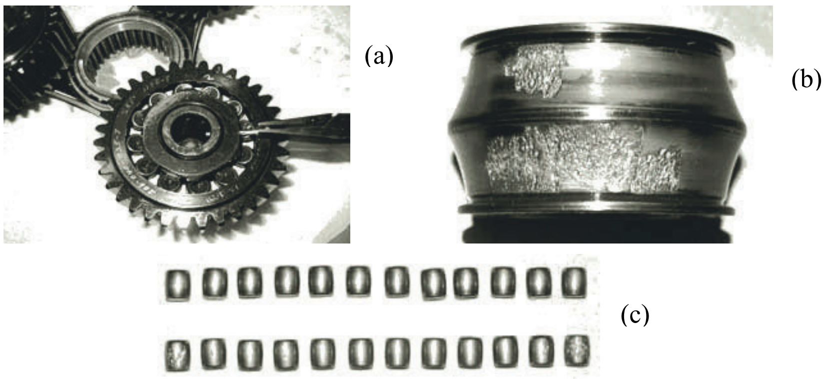


Figure 21 (a) Defective bearing, (b) Spalled inner race (c) Defective rollers

6 CONCLUSIONS

The presented processing algorithm needs minimal user interference to remove the strong gear signal and to determine the suitable band for the envelope analysis. No historical data is required and one sensor would be enough to detect faults in most cases. The presented processing algorithm for detecting faults in rolling element bearings is in fact an easy and effective tool that requires minimal interference and experience from the user. The algorithm is not only able to detect faults in noisy and awkward situations in a clear way but also is capable of tracking the fault severity.

7 REFERENCES

[1] Forrester, BD. (1991) Time-Frequency Domain Analysis Of Helicopter Transmission Vibration, *Department of Defence, Aeronautical Research laboratory*, Propulsion Report 180.

[2] McFadden, PD & Smith, JD. (1985) An explanation for the asymmetry of the modulation sidebands about the tooth meshing frequency in epicyclic gear vibration, *Proceedings of the Institution of Mechanical engineers*, 199(C1), pp65-70,

[3] Lewicki D. & Coy J. (1987) Vibration characteristics of OH58a helicopter main rotor transmission, *NASA technical paper*, NASA TP-2705/AVSCOM TR 86-C-42.

[4] McFadden PD & Howard IM. (1999) The detection of seeded faults in an epicyclic gearbox by signal averaging of the vibration, *Department of Defence, Aeronautical Research lab*oratory, Propulsion Report 183.

[5] Huff E & Tumer I. (2003) Using Triaxial Accelerometer Data For vibration Monitoring Of Helicopter Gearboxes, *Journal of vibration and acoustics*, volume 128 (4), pp120-128.

[6] Dempsey P, Lewicki D. & Decker H. (2004) Transmission bearing Damage Detection Using Decision Fusion Analysis, *NASA technical paper*, NASA/TM-2004-213382.

[7] Ho D & Randall RB. (2000) Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals, *Mechanical Systems and Signal Processing*, 14 (5), 763-788.

[8] Antoni J& Randall RB (2004) Unsupervised noise cancellation for vibration signals: part 2- A Novel Frequency Domain Algorithm, *Mechanical Systems and Signal Processing*, 18 (1), 103-118.

[9] Darlow MS, Badgley RH & Hogg GW. (1974) Application of high frequency resonance techniques for bearing diagnostics in helicopter gearboxes*, Us Army Air Mobility Research and Development Laboratory Technical Report* 74-77.

[10] Bonnardot F, Randall RB & Antoni, J. (2004) Enhanced Unsupervised Noise Cancellation using Angular Resampling for Planetary Bearing Fault Diagnosis, *International journal of acoustics and vibration*, volume 9 (2).

[11] Sawalhi N & Randall RB. (2005) Spectral Kurtosis Optimization for rolling element bearings, *8th International Symposium on Signal Processing and Applications*, Sydney.

[12] Antoni, J. (2006) The Spectral Kurtosis: A useful tool for characterising nonstationary signals, *Mechanical Systems and Signal Processing*, vol 20 (2), pp282-307.

[13] Antoni J & Randall, RB. (2006) The Spectral Kurtosis: Application to the Surveillance and Diagnostics of Rotating Machines", *Mechanical Systems and Signal Processing*, vol 20(2), pp308-331.

[14] Lang G. (1999) S&V Geometry 101, *Sound and vibration*, pp 1-12.

[15] Howard I. (1994) A review of rolling element Bearing Vibration: Detection, Diagnosis and Prognosis, *DSTO-AMRL report*, DSTO-RR-00113, pp.35-41.

[16] Wang W and Wong AK. (2002), Autoregressive Model-Based Gear Fault Diagnosis, *Transaction of ASME, Journal of Vibration and Acoustics*, Vol. 124, pp 172- 179.

[17] Sawalhi N & Randall RB. (2005) Spectral Kurtosis Enhancement using Autoregressive Models. *The 4th Australasian Congress on Applied Mechanics*, Melbourne.

[18] Akaike H. (1969) fitting autoregressive for prediction, *Ann.Inst.Math*, vol (21), pp243-247.

ACKNOWLEDGMENTS

This work is supported by the Australian Defence Science and Technology Organization (DSTO) as a part of their Centre of Expertise scheme.

DEVELOPMENT AND IMPLEMENTATION OF MAINTENANCE PERFORMANCE MEASUREMENT SYSTEM: ISSUES AND CHALLENGES

Uday Kumar

Division of Operation and Maintenance Engineering, Luleå University of TechnologyV971 87 Luleå, Sweden

Abstract: The cost of maintenance and its influence on the total system effectiveness is too great for asset managers to ignore. The modern days asset managers need to know the relationship between the outputs from the investments in the maintenance in terms of total contribution to the business goal and the inputs to maintenance processes. In addition, without any formal measures of performance, it is difficult to plan control and improve the maintenance process. The discussion in the paper is based on some of the R &D projects being pursued at the Division of Operation and Maintenance Engineering at Luleå University of Technology, Luleå in close cooperation with infrastructure owners and asset managers in Sweden. This paper will discuss the issues and challenges associated with development, implementation of maintenance performance system and maintenance performance indicators for effective management of infrastructure and industrial assets.

Key Words: Maintenance performance, Total maintenance effectiveness, Overall equipment effectiveness (OEE), Maintenance performance indicators, external effectiveness, internal effectiveness.

1 BACKGROUND AND INTRODUCTION

Many companies with capital intensive assets have accepted that an effective management of maintenance process is key to sustaining the long term profitability of their business organizations. With increasing awareness that maintenance creates additional value in the business process; more and more companies are treating maintenance as an integral part of the business process, and the measurement of maintenance performance has become an essential element of strategic thinking of many companies involved in service and manufacturing industry. Without any formal measures of performance, it is difficult to plan control and improve the maintenance process. Maintenance as an important support function in business with significant investment in physical assets plays an important role to achieve the organizational goals [1]. The following factors have made pressing demands towards an effective management of maintenance performance even more challenging:

1.1 Measuring value created by the maintenance

The most important reason for implementing maintenance performance system is to measure the value created by maintenance process. As a manager, one must know that what is being done is what is needed by the business process, and if the maintenance output is not contributing/creating any value for the business, it needs to be restructured. This brings the focus on doing right things keeping in view the business goal of the company [2].

1.2 Justifying investment

The second basic reason for measuring maintenance effectiveness is to justify the organization's investment made in maintenance organisation; not so much as to whether you are doing the right thing, but whether the investment they are making is producing a return on the resources that are being consumed.

1.3 Revising resource allocations

The third basic purpose for measures of effectiveness is to determine if additional investment is required and to justify the investment. Alternatively, such measurement of activities also permits you to determine whether you need to change what you are doing or how you can do it more effectively by using the resources allocated.

1.4 Health safety and environmental (HSE) issues

The fourth reason can be to understand the contribution of maintenance towards HSE issues. A bad maintenance performance can lead to accidents (safety issue) and pollutions (health hazards and environmental issues), besides encouraging an unhealthy work culture and environment.

1.5 Focus on knowledge management

Many companies especially those involved in delivery of maintenance and product support services are focused on effective management of knowledge in their companies. Furthermore**,** Technology is ever changing and is changing faster in the new millennium. This has brought in new sensors and embedded technology, information and communication technology (ICT) and condition based inspection technology like vibration, spectroscopy, thermography and others, which is replacing preventive maintenance with predictive maintenance. This necessitates a systematic approach for the knowledge growth in the field of specialization.

In the following sections, an attempt has been made on the basis of our experiences gained from R & D projects in the area of performance measurements from Norwegian Oil and Gas industry [3, 4], Swedish Mining Industry [5] and Swedish Railway sector [6, 7] and from literature survey to discuss these issues and challenges associated with development and implementation of maintenance performance indicators (MPIs).

The objective of this paper is to examine and discuss some of the important the issues and challenges associated with development and implementation of a maintenance performance system.

2 MAINTENANCE PERFORMANCE SYSTEM INDICATORS

Maintenance performance measurement is a complex task involving measurement of multiple inputs and multiple outputs. One way of measuring the performance is to develop performance indicators and implement them with a total involvement of entire organisation. It's more or less similar to implementing Total Productive Maintenance (TPM) [8]. Maintenance performance indicators (MPIs) are used for the measurement of maintenance performance, as performance indicators are an indication of performance [9]. Also, an indicator is a product of several metrics (measure), when used for measurement of maintenance is called a MPI. Maintenance Performance Indicators (MPIs) show the value of the contribution of maintenance process towards the ultimate objective of the company as a whole. The structure of the MPIs needs to be considered from different aspects, like the multi-criterion and hierarchical levels and from total maintenance effectiveness.

Measurement tells us the status of the job carried out and what action to be taken there after, and to indicate where those actions should be targeted. The selection of MPI depends on the way in which the maintenance performance measurement (MPM) system is designed and developed for application. For example, the MPIs could be used for financial reports, for control of performance of employees and other resources like the costing and appraisal system, for finding competitive position with in business organizations like the customer satisfaction and competitor ranking, for health, safety and environmental (HSE) rating for production industry, and finding internal effectiveness, like the overall equipment effectiveness (OEE) for the manufacturing and process industry. The selection of maintenance performance indicators to follow up the contribution of maintenance is an important but a complex issue. The structure of the performance indicator needs to be considered from different aspects.

SMART test developed by the Department of Energy (DOE) can be used effectively to test the attributes of the indicators [10]:

S: Specific; clear and focused to avoid misinterpretation. Should include measure assumptions and definitions and be easily interpreted. Maintenance cost/ton of ore

M: Measurable; can be quantified and compared to other data. It should allow for meaningful statistical analysis. Avoid "yes/no" measures except in limited cases, such as start-up or systems-in-place situations.

A: Attainable; achievable, reasonable, and credible under conditions expected.

R: Realistic; fits into the organization's constraints and is cost-effective.

T: Timely, the indicator should be reflecting the status in real time and on time.

3 TOTAL MAINTENANCE EFFECTIVENESS

A quick search of literature shows no dearth of performance measures being used by industry. Unfortunately most of them are focused on either measuring financial results or the health of the process in piece meal manner such as system availability, consumption of resources per unit of measures adopted etc. It was difficult to find any discussion on measuring total effectiveness of maintenance process. Some of the notable exceptions are from Sweden [14].

The total effectiveness is a product of the internal effectiveness characterised by issues related to effective and efficient use of resources to facilitate the delivery of the maintenance and related services to be delivered in the most effective way (engineering and business process related to planning and resource utilization) and external effectiveness characterised by customer satisfaction, growth in market share etc. [16]. See figure 1.

The performance measures for internal effectiveness are concerned with doing things in right way and can be measured in terms of cost effectiveness (maintenance costs per unit produced), productivity (number of work orders completed per unit time) etc. and deals with managing resources to produce services as per specifications.

The performance measures for external effectiveness deals with measures that have long term effect on companies profitability and are characterised by doing right things that is delivering services in a way (quality and timeliness) that as per customer requirement. Here the concept of delivering not only the services required by customers; but helping them in their other business process related to their own services. Such attitude often helps in market growth and capturing or creating new markets.

Whenever, we develop a balanced maintenance measurement system, we have to examine all the related criteria and parameters associated with the system. The concept of total effectiveness includes both the external and internal effectiveness. In any organization, first the maintenance process needs to be studied in detail and external effectiveness factors like the stakeholders requirements (front end processes) need to be understood. Then based on the internal resources and capabilities, supply chain management (back end processes), the maintenance objectives and strategies are formulated matching and integrating with that of the corporate ones. An important objective of the measurement system should be to bridge the gap and establish the relationship between the internal measures (causes) and the external measures (effects) [17].

Traditionally, the concept of OEE [8 used by the company is inadequate as it only measures the internal effectiveness. For example, the OEE level of an organization may be at 80%, but if the external effectiveness is as low as 60 %, the total maintenance effectiveness will certainly be low. The index for customer satisfaction; like the, quality (of repair/modification), timeliness of delivery, health, safety and environmental (HSE) index, and call back of product (promptness of response), are the decisive factors of total maintenance effectiveness as given at Figure 1.

For measuring the total maintenance effectiveness, a balanced, multi-criteria and hierarchical MPM system is considered to be effective, which considers both the external and internal effectiveness [18]. A need is also exists to workout an overall total maintenance effectiveness considering all the factors and criteria as discussed above.

In general measures for total maintenance effectiveness must be combined with process owners' capability to change processes and adopt to new technology and work practices without any major involvement of resources and at right time.

Front end processes

Total maintenance effectiveness	External Effectiveness	1. Customer satisfaction index - Quality - Timeliness of delivery
	Internal	Productivity Cost per unit

Back end processes

Total maintenance effectiveness = Internal effectiveness x External effectiveness

Figure 1. Total maintenance effectiveness

The MPIs are also considered from multi hierarchical levels of the organization. From hierarchical levels, the first layer could be at the corporate or strategic level, the second level deals with the tactical or managerial level and the third level deals with the functional/operational level. Depending on the organizational structure, the hierarchical levels could be more than three. Three level hierarchical levels are given in Figure 2.

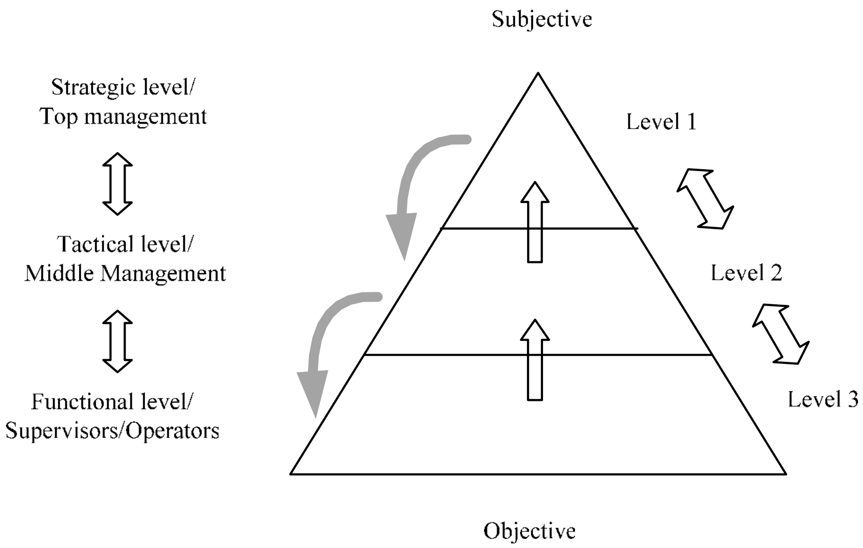


Figure 2. Subjectivity increases as we integrate the objective outcomes fromthe shopfloor to get key performance indicator at higher level

4 ISSUES AND CHALLENGES IN DEVELOPING MPIS

There are numerous issues and challenges associated with development of maintenance performance indicators which needs to be considered while developing and implementing MPIs in a company. Some of these are "what should be measured and how many measures are required, what feedback is necessary and who should receive the details of measures, what type of data is required and how to collect data, how the data is going to be used and analysed (which information software), what kind of information is needed for decision making, who should be responsible for measurement and reporting, and, who should sponsor/own the overall measurement process?. Many of these are taken care of by various maintenance indicators developed during the last 2-3 decades. Most of these are focused on measuring the internal effectiveness [1, 11, and 12]. Some of the researchers have focussed on external effectiveness with special focus on external stake holders requirements [13].

However the main issues of measuring the total maintenance effectiveness is still under study by many researchers involved in studying the maintenance business process and its work processes [14, 15].

The major issues challenges while developing MPI are namely:

- Measurement of Total maintenance effectiveness (TME): How to develop measures that reflect the total maintenance effectiveness?
- Involvement of strategic and operating managers in developing and implementing performance measures; How to involve every one in the process of maintenance performance measurement? Involvement of all employees in MPIs development and implementation process, so that everyone speaks the same language. As recommended in TPM philosophy, the measurement process begins at the highest levels within companies.
- Corporate goal and vision: How to integrate and link the outcome from maintenance at shop floor level into the corporate business goal and strategy at top management level? How to break down the corporate goals and strategy into more objectives targets/tasks for operating managers and engineers at shop floor level and integrating and linking the objective output at shop floor to the corporate goals and strategy? See for details [15].

Before commencement of the development process for indicators, a detailed understanding of the maintenance business process and its strategic positioning in the company is a must.

It is a challenge to inetgrate MPIs generated at shop floorlevel by operating managers to key performance indicators and linking back them to companies strategic visions and goals. The following questions needs to be answered to determine the effectiveness and quality of the MPI; those are identified to be used in the MPM system.

4.1 Does the MPI take care of stakeholders requirements?

The MPM system has to take care of the stakeholders' requirements from the organizational effectiveness, both from external and internal perspectives. In fact this should be the prime requirement of the MPM system.

4.2 Is the MPI measurable objectively?

The MPI should be measurable in comparison with the set target. It should be quantifiable and indicate an objective result of the maintenance performance. The indicator should be specific and must be identifiable with the objective it is meant for.

4.3 Does the MPI looks in to the efficiency and effectiveness of the organization?

The organizational effectiveness and efficiency are most important issue in to day's very dynamic competitive business. The importance of the effectiveness and efficiency of the operation and maintenance process of the organization needs no emphasis. The MPI formulation therefore, must look in to these aspects from the conceptual stage.

4.4 Does the MPI indicate a clear statement of the expected performance results?

The MPI need to be specific and clearly state the performance it is going to measure.

4.5 Are the MPIs challenging and yet attainable?

The MPIs should be achievable and at the same time they should be challenging. They should be able to create a interest for continuous improvement amongst the employees of the organization.

4.6 Are the MPIs linked to the benchmarks or milestones quantitatively /qualitatively?

The MPIs should be linked to the benchmark or standardization strategically, so that increasingly achievable milestones can be set. These milestones are set keeping the capability, resources and plant capacity in mind.

Some of the questions (5Ws and 1H) that need to be considered/answered, while developing the performance measurement system and indicators are:

- why measurement is needed,
- what should be measured to provide the right information and what type of data is required,
- who should own the process and measure it,
- where the details of measures are required to be undertaken
- when the data and information are required for decision making
- how to collect the data, and making it clear to all involved that, data is neither information nor knowledge, but provides a basis for making correct decision about the status/health of the system

Finding answers to these questions will lead to development of more relevant maintenance indicators and will also facilitate easy and effective implementation of the performance measuraement system in the company.

5 SUMMARIES OF THIS PAPER AND SCOPE FOR FUTURE WORK

In this paper some of the most important maintenance performance measurement issues and challenges are presented in brief. It is also proposed that a multicriterion hierarchical MPIs and MPM system be developed for a relevant, interpretable, reliable, cost and time effective, easy to implement, update and maintain the system for regular use by stake holders at various levels. The indicators at subsystem/ component level, plant level and corporate level are linked with that of the organizational objectives and strategy. Each measure that is selected must be linked back to the company's vision and goals. If this cannot be done, then the measure does not have the requisite attributes. The first step in developing MPIs is to involve the maintenance crew, who are responsible for the work and understanding the maintenance and business work process.

6 REFERENCES

[1] Tsang A.H.C. (2002) Strategic dimensions of maintenance management, *Journal of Quality in Maintenance Engineering*, Vol. 8, No. 1, pp. 7-39

[2] Herring J.P. (1999) Measuring Effective Intelligence: Meeting the Management Communication Challenge, http://www.cireport.com/herring/herring_1.html, Dated 2008-2-06

[3] Kumar U & Ellingsen H.P. (2000) Development and implementation of maintenance performance indicators for the Norwegian oil and gas industry. *Proceedings of the 14th International Maintenance Congress* (Euro-maintenance, 2000), 7-10 March 2000, Gothenburg, Sweden, pp. 221-228.

[4] Ellingsen H.P, Kumar U *et. Al.* (2002) Management of assets, resources and maintenance by using a balanced scorecard based performance framework, *Proceedings of the 16th International Maintenance Conference*, (Euro maintenance 2002), pp 203-211

[5] Parida A & Nissen A. (2006) Development of Maintenance performance Indicators for a Pelletization Plant at LKAB, Internal Report, Division of Operation and Maintenance Engineering, Luleå university of technology, Sweden.

[6] Åhren T. (2005) A study of Maintenance Performance Indicators for the Swedish Railroad System , Licentiate Thesis: 2005:47/ISBN: 1402-1757, Luleå University of Technology, Sweden

[7] Åhren T & Kumar U. (2004) Use of maintenance performance indicators: a case study at Banverket. *Conference proceeding of the 5th Asia-Pacific Industrial Engineering and Management Systems Conference* (APIEMS2004), Gold Coast, Australia

[8] Nakajima S, (1988) Introduction to TPM: Total Productive Maintenance, Productivity Press

[9] Wireman T. (1998) Developing Performance Indicators for Managing Maintenance, New York, Industrial Press, Inc.

[10] DOE-HDBK-1148-2002 (2002) Work Smart Standard (WSS) Users' Handbook, Department of Energy, USA, www.eh.doe.govt/tecgstds/standard/hdbk1148/hdbk11482002.pdf

[11] Al-Turki U & Duffuaa S. (2003) Performance measures for academic departments, *International Journal of Educational Management,* Vol. 17, No. 7, pp. 330-338

[12] Coetzee J.L. (1997) Towards a general maintenance model, Martin, H. H. (ed), New development in maintenance *proceedings of the International Federation of Research in Maintenance Conference* , pp 1-9

[13] Liyanage (2003) PhD Thesis, Operation and Maintenance Performance in Oil and Gas Production Assets, Norwegian University of Science and Technology, Trondheim, Norway.

[14] Ahlmann H. (1999) The economic significance of maintenance in industrial enterprises, UTEK, (Swedish Maintenance Society) , Seminar

[15] Parida A & Kumar U. (2005) Maintenance performance measurement (MPM): Issues and challenges, Submitted for publication to journal of Quality in Maintenance Engineering.

[16] Bruzelius L.H & Skärvad P-H (2004) Integreraded Organisationslära (In Swedish), Studentlitteratur, Lund

[17] Jonsson P & Lesshammar M. (1999) Evaluation and improvement of manufacturing performance measurement systems: the role of OEE, *International Journal of Operations & Production Management*, Vol. 19, No. 1, pp. 55-78

[18] Parida A, Chattopadhyay G & Kumar U. (2005) Multi criteria maintenance performance measurement: a conceptual model, *Proceedings of the 18th International Congress COMADEM*, 31st Aug-2nd Sep 2005, Cranfield, UK, pp. 349-356

Acknowledgement:

The author would like to thank the colleagues at the Division of Operation and maintenance especially Aditya Parida, lecturer for their contribution in preparation of this manuscript.

DESIGN AN INTEGRATED INTELLIGENT EQUIPMENT MAINTENANCE AND PRODUCTION DISPATCHING SYSTEM USING THE JESS PLATFORM

AJC Trappey C Trappey, GY. Lin PS Ho CC Ku CL Kuo

[a] *Dept of Industrial Engineering & Engineering Management, National Tsing Hua University, Hsinchu 300, Taiwan*

[b] *Queensland University of Technology, Australia.*

[c] *Avectec.com, Inc., 3F-2, 29 Puding Road, Hsinchu 300, Taiwan*

[d] *Mechanical Industry Research Laboratory, Industrial Technology Research Institute, Chutung, Hsinchu 310, Taiwan*

Abstract: In modern factories, production equipment maintenance represents a significant and important task to ensure that plant facilities continue to function properly. On the other hand, intelligent production dispatching generates maximal outputs for changing production demands. In this research, a system architecture, called Integrated Intelligent Equipment Maintenance and Production Dispatching (IEMPD) building on the JAVA Expert System Shell (JESS), is designed and developed to support collaborative maintenance, remote equipment diagnosis, job dispatching, and production simulation/verification. The IEMPD system is applied for dynamic shop floor equipment maintenance and production control. In the past, trouble shooting analyses were based on the experiences and skills of equipment engineers. The IEMPD platform allows knowledge engineers to construct, manage and maintain rules easily and flexibly to satisfy ever-changing collaborative maintenance and production requirements. Through the rule base, maintenance knowledge and experiences can be created, accumulated, shared and reused. Thus, IEMPD can rapidly help maintenance engineers identify the faults and suggest solutions. Further more, IEMPD can assist production engineers plan dispatched jobs based on the orders and the equipment status. This research uses a machining flexible manufacturing system (FMS) as the case study to demonstrate the IEMPD capability in supporting integrated intelligent maintenance and production.

Key Words: Rule Base, Knowledge System, Expert System, Intelligent Equipment Maintenance System, Production Dispatching, JESS

1 INTRODUCTION

Along with the global changes of market conditions as well as the impact of competition, speed and flexibility of production become key success factors for enterprises. One who can produce goods according to customer's demand, shorten delivery time and offer products with high quality and good customer services will survive the harsh challenges in the global market-place. Nowadays, factories are capital-intensive requiring highly sophisticated and automated production equipments. Therefore, the most challenging task during manufacturing processes is to diagnose and analyze the equipment problems and identify proper maintenance and repair methods effectively. Research has shown that a company can easily increase 20% of its operating cost if it does not maintain its production equipment properly [2]. Hence, it is important to develop intelligent diagnosis and maintenance systems to prevent equipment failures and unexpected breakdowns. An intelligent diagnosis and maintenance system should be Internet/web-enabled. It consists of intelligent machine monitoring, e-prognostic and e-diagnostics to keep manufacturing operation machines in good operating condition. Besides effective monitoring of the machine statuses and the good diagnosis and maintenance analysis to ensure that equipment continues to function properly, it is important to make the operation processes stable and smooth according to shop floor control logics. There is a growing interests in intelligent manufacturing shop floor control, specifically in the area of simulation analysis to predict the future impact of short-term decisions on the performance of a manufacturing system. With on-line simulation, the current status information can be used to predict system performance and to develop short-term scheduling and control alternatives. In order to perform the above-mentioned various tasks, an intelligent system must collect, extract, and manage adequate domain knowledge constantly [23]. Therefore, there are many research focusing on improving the rule-based expert systems. Lin, Tseng, and Tsai (2003) [15] proposed an object-oriented rule-based paradigms knowledge model based on the concept of learning and thinking behaviors of human. A rule-based paradigm provides us with general facilities for deductive retrieval and pattern matching, and the object-oriented programming provides us with an intuitive structure for programming in the form of class hierarchies.

In this research, an Integrated Intelligent Equipment Maintenance and Production Dispatching (IEMPD) platform using web-enabled JESS technology is designed and developed to demonstrate how the tasks can be realized and implemented effectively in the FMS maintenance domain. Further, this research uses the FMS dispatching control case to demonstrate the knowledge-based reasoning capability for intelligent production planning.

This paper is divided into five sections. In Section 2, the background research for rule-based expert system, JESS technology, and intelligent diagnosis and maintenance system are depicted. In Section 3, we describe the conceptual architecture of IEMPD system. Section 4 reveals a case study for the shop floor dispatching control of a machine shop FMS. Finally, the conclusions are drawn in Section 5.

2 BACKGROUND RESEARCH

In this section, the background research is divided into three parts detailing rule-based expert system, JESS technology, and intelligent maintenance and production planning issues.

2.1 Rule-based Expert System

The artificial intelligence (AI), since its inception in 1960's, has been widely developed in a variety of domains to solve problems and support decision making. Among well established AI methodologies, the most researched area is in Expert System (ES) [12] [21] [28]. Expert system uses heuristic knowledge captured in a software program to solve problems that ordinarily require human reasoning and accumulated expertise [24]. Figure 1 shows the typical architecture common to rule-based expert system (RBES), consisting of Knowledge Base (Rule Base), Fact Base (Working Memory), and Inference Mechanism (Rule Engine).

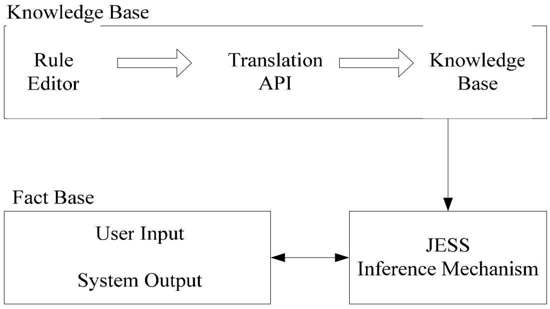


Figure 1. The architecture of a rule-based expert system [30]

- Knowledge Base (Rule Base): The development of an expert system involves the construction of the knowledge base by acquiring knowledge from experts or documented sources. The knowledge base contains domain knowledge useful for problem solving in rule formation [20]. Rule is a natural knowledge representation, in the form of the 'IF … Then…' structure.
- Fact Base (Working Memory): The facts database contains a set of facts, which can be entered by user or created automatically by the system. Facts must be presented in the working memory in order for a rule to become available for activation [17]. A set of facts and a set of rules that can be fired by patterns form the assertions. The basic operations of the assertions are to manipulate facts and rules in the knowledge domain, such as defining facts and rules, deleting rules, creating rule sets, as well as adding and removing facts from working memory.
- Inference Mechanism (Rule Engine): The inference mechanism, an interpreter of rule-based expert system, uses an iterative match-select-act cycling model. In act phase of the cycle, a fired rule may modify or generate some facts [15].

An rule-based expert system (RBES) has many advantages [19]. The first advantage is naturalness of expression since experts rely on rules rather than on textbook knowledge. The second advantage is modularity that permits systems easy to construct, to debug, and to maintain. Restricted syntax and ability of explanation are also the advantages of RBES. RBES has been widely applied in several domains. For instance, Wagner et al. (2002) [27] has proposed using a knowledge-based system (KBS) to financial problem domains. Yao, Lin, and Trappey [29] have used knowledge-based intelligent reasoning to support dynamic equipment diagnosis and maintenance. In the production verification aspect, Chiang, Trappey and Ku have developed a mechanical design intelligent system using JESS knowledge platform for collaborative designs [7].

2.2 JESS Technology

Built on top of the Java language platform, Java Expert System Shell (JESS) has become a popular ES development tool for intelligent system development. JESS is a simple and powerful tool to allow for the building of multiple ES applications in a single platform. JESS is essentially a reimplementation of a subset of the earlier CLIPS shell [8] in Java environment. CLIPS is the original expert system shell that formed the inspiration for Jess. Jess uses a customized Rete inference engine that is different than the one in CLIPS, but Jess can run most CLIPS files.

The JESS shell provides the basic elements of an expert system including fact-list, knowledge base that contains all the rules and inference engine which controls overall execution of rules as shown in Figure 2. Figure 2 shows the architecture of JESS and how all the parts fit together, including:

- Pattern Matcher: decides what rules to fire and when.
- Agenda: schedules the order in which activated rules will fire.
- Execution Engine: responsible for firing rules and executing other code.

When executing inference process, rules and facts are compared using pattern matcher. The matched rules are activated into a conflict set. Afterward, the conflict set resolves into agenda (called conflict resolution) and, then the execution engine fires on the agenda. The engine fires only once on any given set of facts in the working memory. The engine cycles recursively until all rules are satisfied.

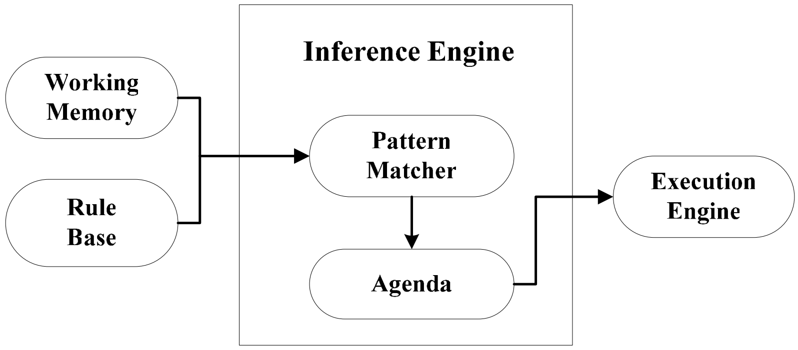


Figure 2. The architecture of JESS [10]

There are three types of facts defined in JESS [11]:

- Ordered Facts: Ordered facts are simple lists, where the first field (the head of the list) acts as a sort of category for the fact.
- Unordered Facts: Ordered facts are useful, but they are unstructured. Sometimes there is the need for organization. In the object-oriented languages, objects have named fields or slots in which data appears. Unordered facts offer this capability.
- Shadow Facts (Definstance Facts): Shadow facts are unordered facts whose slots correspond to the properties of a JavaBean. JESS lets user put Java objects into working memory as long as they follow the JavaBean specification. Shadow facts connect JESS' working memory to the application containing the JESS engine.

The expert knowledge and user inputs are stored as rules and facts in the JESS knowledge engine. The rules are "if then" statements that have a left-hand-side part (IF-portion) and a right-hand-side part (THEN-portion) with the "implies" operator in between. Rules are responsible for taking actions based on facts in the knowledge base. Whenever the patterns of the rule match with existing facts, the rule is activated and its right-hand-side actions executed. JESS uses the Rete algorithm to match the rules and facts. Then the rules will be executed to produce different results accordingly [4]. JESS's major advantage is its capability to easily integrate with other Java programs through its well-defined API [9].

2.3 Intelligent Maintenance and Production Planning

The development of on-line equipment diagnosis and maintenance systems, based on the latest information technology, has gained popularity among equipment providers and users [3] [25]. Some literatures [14] [6] proposed SPC and Run-by-Run methods to control the quality and eliminate the variations. Those methods not only show the quality of products but also reflect the running condition/status of the working equipments. Many papers have attempted to combine two or more methods, e.g., neural network, expert system and fuzzy system, to monitor or diagnose manufacturing systems. For example, Wang, Liu, and Griffin (2000) [26] developed a combination system of neural network and expert system for diagnosis with superior performance comparing to the old approaches. Shal and Morris (2000) [22] proposed a fuzzy expert system for the equipment fault detection using statistical process control of industrial processes.

Preventive maintenance affects the performance of equipments significantly. Efficient maintenance increases productivity and production quality, as well as reduces product defects. Some research have developed remote maintenance control through Internet. Lee (2001) [13] suggested a framework for web-enabled and Internet-based Intelligent Maintenance System (IMS) to seek near-zero-breakdown equipment conditions. Balakrishnan and Honavar (2003) [1] examined and compared several different approaches to the design of intelligent systems for diagnoses. Bengtsson (2003) [2] discusses the technical components of a complete condition-based maintenance system for industrial robot fault detection and diagnosis. About production plan and performance evaluation, Ngai and Cheng (2001) [16] proposed expert system application for performance measurement of advanced manufacturing technology projects. Performance evaluation or productivity analysis also fall in the area of auditing and internal control assessment [5]. Rao, Miller, and Lin (2005) [18] describes the framework for a comprehensive productivity analysis procedure and presents a prototype expert system that attempts to provide this analysis.

Although many systems and tools have been developed in the maintenance and production domains, effectively and dynamically designing, implementing and updating these systems, especially in e-manufacturing maintenance applications, is still a challenging task. The challenges are mainly in the ill-defined nature of the fault symptoms, real-time discovering of faults, complexity of knowledge representation and ever-updating diagnosis and maintaining rules. This paper addresses how the maintenance and production knowledge can be expressed dynamically and utilized in various sites using cyber-enabled JESS technology.

3 CONCEPTUAL ARCHITECTURE OF IEMPD

In this research, a web-based prototype IEMPD, which combines rule-based knowledge system and inference engine, is designed and implemented to support collaborative maintenance, remote equipment diagnosis, job dispatching, and related production decision via Internet. The system architecture and main functions of IEMPD, conceptual modeling of knowledge base and rule inference process are presented in the following sub-sections.

3.1 System Architecture and Main Functions of IEMPD

Figure 3 shows the conceptual architecture of the Integrated Intelligent Equipment Maintenance and Production Dispatching (IEMPD) platform. The system architecture is divided into three parts (i.e., IEMPD, User, and Data) to ensure the flexibility of system. All messages in the platform can be transmitted fast and efficiently through the Internet. IEMPD consists of three main system modules, i.e., (1) Knowledge Management Module, (2) Project Management Module, and (3) System Administration Module. The main functions of these modules are outlined as follows:

1. Knowledge Management Module has the following sub-functions.
   - Knowledge Type Management
   - Knowledge Component Management
   - Knowledge Component Attribute Management
   - Document Management
   - Mathematical Function Definition
   - Mathematical Function Attribute Management
   - Measurement Unit Management
   - Measurement Unit Conversion Management
   - Rule Management
   - Rule Sets Management
2. Project Management Module consists of sub-functions as follows.
   - Project (Case) Management
   - Project Role Management
   - Project Functions Management
3. System Administration Module includes sub-functions as follows.
   - System Administration
   - Authority Management
   - New Functions Suggestion Management
   - System Functions Management
   - User Management
   - Role Management

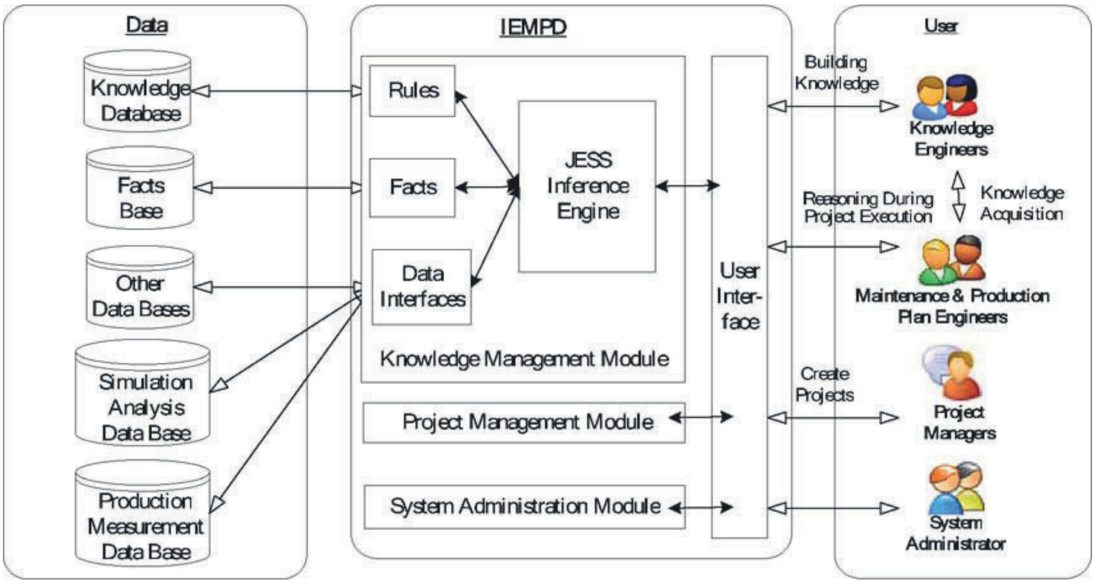


Figure 3. The conceptual architecture of IEMPD

IEMPD allows different members of teams participate through Internet for collaborative maintenance, remote equipment diagnosis, job dispatching, and production simulation/verification. IEMPD has four roles of users participating collaboratively in the system platform. They are knowledge engineers, project manager, maintenance and production plan engineers, and system administrators.

Knowledge engineers input knowledge rules into project database and fact database through the interface of knowledge rule management, which is responsible for managing and maintaining the entire knowledge base. A project manager (PM) can create, execute, save, and delete projects. He/She also has the authority to monitor and manage projects belonging to him/her. Still, a PM can define related variables and select the needed knowledge rule modules. Maintenance and production plan engineers have authority to execute tasks and modify their own profiles. They can input necessary parameters for the reasoning of optimal maintenance or production strategies when execute project tasks. Knowledge engineers usually acquire experiences and knowledge of the production/maintenance from senior equipment maintenance and/or production engineers. A system administrator has the highest authority to permit system functions to specific users. Figure 4 shows a use case diagram in Unified Modeling Language (UML) format for an overview of the IEMPD system modules.

In the data section, project database saves the data about maintenance or production projects (including the definition of projects, tasks, and process logics). The project database will produce rules into inference engine. Fact base saves the fact-format data. When facts in inference engine change, data in fact database will also be updated. IEMPD will connect other databases (e.g., simulation analysis database and production measurement database) to enhance the system capability.

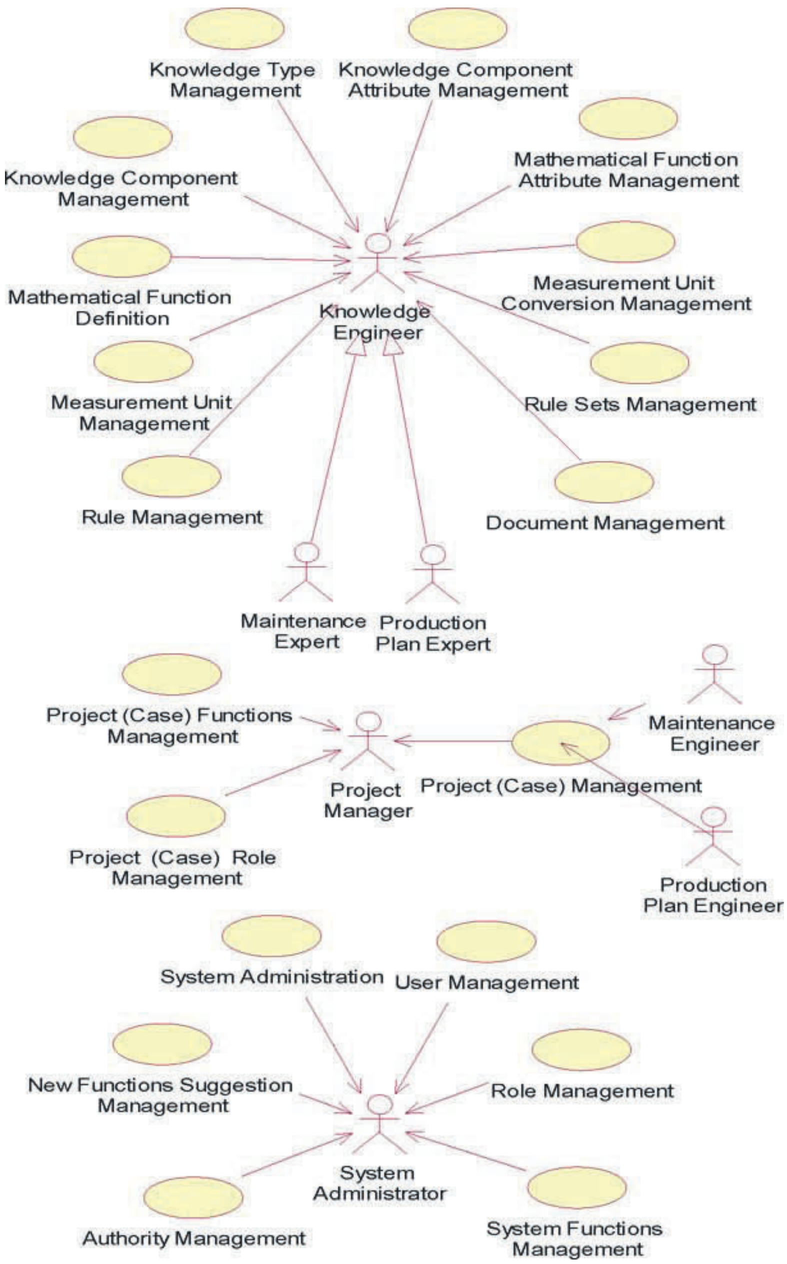


Figure 4. The UML use case diagram for an overview of system functions

3.2 Knowledge Management Module and Its Rule Inference Process

The detailed elements of the knowledge management module are shown in Figure 5. They include the definition and management of knowledge types, knowledge components, mathematical models, knowledge component attributes, mathematical model attributes, documents, measurement units, and rules.

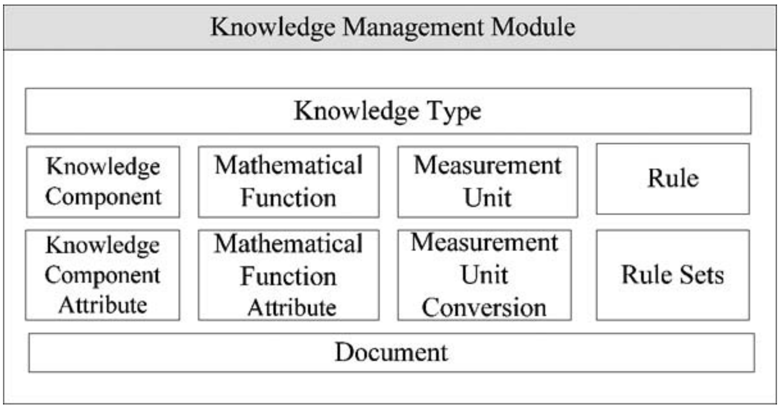


Figure 5. The detailed elements of the knowledge management module in IEMPD

When executing a project for maintenance or production decision, it is important to realize and analysis the relationship between the possible cause and breakdown indications for accurate diagnoses. To detect problems effectively, it is necessary to divide the machine or system into many modules and sub-modules that contain individual rules and facts for inference.

Based on a cognitive analysis of maintenance tasks, the rules are developed. Thus, maintenance and production plan engineers can make decision immediately by using IEMPD inference. In the rule inference process of the IEMPD, some basic data, such as knowledge types, knowledge components, mathematical models, and domain knowledge rules must be defined prior enabling and activation of all functions and rules. The flow diagram in Figure 6 shows the steps of utilizing IEMPD for every case/application of equipment maintenance and supporting production plan.

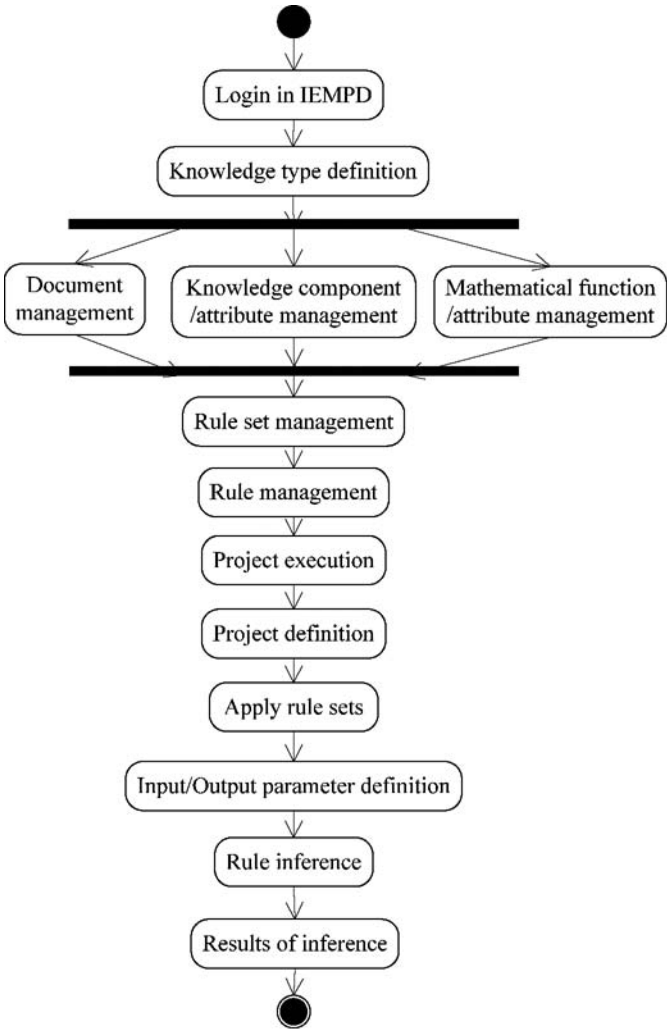


Figure 6. The rule inference process of the IEMPD

4 CASE STUDY

In this research, a prototype IEMPD is developed, which combines rule-based knowledge system and inference engine, to support job dispatching decision in the manufacturing environment. A machining shop FMS dispatching control project is used as a case study. FMS is a flexible and reliable automation system with huge productivity. The FMS consists of various types of capital equipments, e.g., Computer Numerical Controlled (CNC) machines and automated guided vehicles (AGVs). In FMS, the central control computer carries out the flow of works and some of its functions include scheduling jobs onto the machine tools, downloading CNC part-programs, and sending instructions to the AGV system for transportation. By using FMS, manufacturers can produce customized workpieces at a faster rate, making it easier to meet market demand. In the case study, an example FMS machine shop is used to describe the functions of the IEMPD (as shown in Figure 7). The system consists of three CNC milling machines, one CNC turning machine, two industrial robots, inspection station, an automated part storage retrieval system (AS/RS), an AGV for transportation, and an automated cart-based conveyor system. In addition to the automated equipments, a human operator is used to load part onto the AGV.

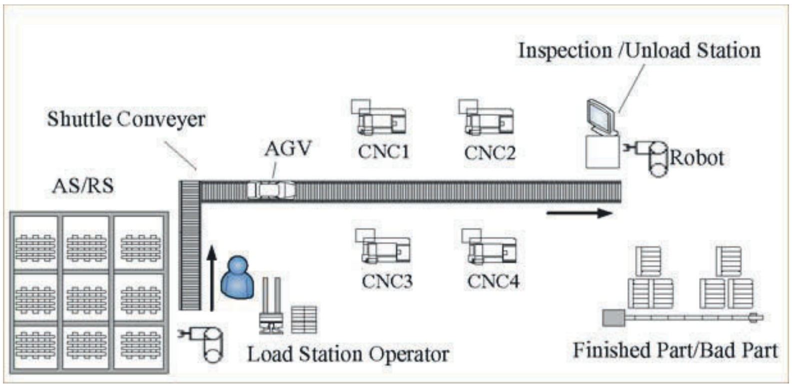


Figure 7. Layout of the machine shop FMS example (NTUT, 2005)

There are nine pallets which are deposited in the AS/RS. Products to be produced are manually loaded onto pallets at a loading station, and the computer system takes over, moving the parts to the various processing stations using AGV. If the next processing station is busy, the AGV will move the part and pallet to the AS/RS according to the Pallet-Number and wait for the next processing station to be idle. After having visited all necessary stations, the part is taken to the unloading station where it is inspected before removing from the pallet by robot. Each processing stations (i.e. CNC1-CNC4, load, unload) has a part buffer for the waiting-in-process (WIP) queues. This research uses the shop floor dispatching control of the example FMS to demonstrate the functions of IEMPD. This integrated IEMPD and FMS structure drives the JESS rule-engine as a “dispatcher” that handles the actual system shop floor control. The related information, e.g., equipment status and part status, can be collected and acquired via monitoring PLC or RFID reader of the FMS as shown in Table 1.

Table 1

Messages sent within the system

Concerning Equipment Information	Concerning Part Information
Operation Machine Status	Part Number and Status
AGV Status	Pallet Number Status
Success Signal of the Operation	Due Date
	Processing Times
	Moving or Waiting
	Processing Routing

Figure 8 shows the data flow through the integrated IEMPD and FMS structure. There is a timer application, which is programmed by the Java Servlet to scan the tables in the intermediary database. Therefore, the information can be trigger and passed to the JESS (Java Expert System Shell) for inferring. The dispatching rules include Shortest processing time (SPT), Earliest Due Date (EDD), and First-in-First-out (FIFO). They are formulated and stored in the fact base for applying to specific project. When IEMPD detects any dispatching indications, JESS automatically generates further message about the next AGV moving guide (from-to). Figure 9 shows the dispatching inference process of the integrated IEMPD and FMS structure using unified modelling language (UML). In this case, the situations of the manufacturing status for the FMS shop floor dispatching decision control are shown in Table 2, Table 3 and Table 4.

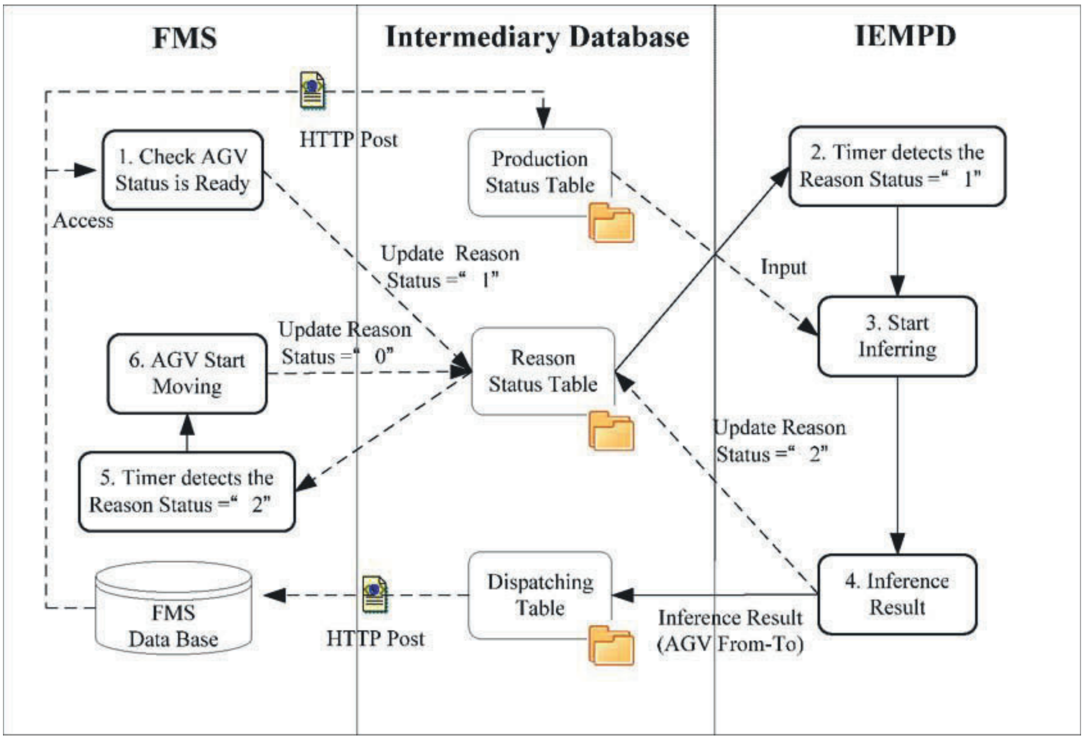


Figure 8. Data flow through the integrated IEMPD and FMS structure.

Table 2

Illustrated examples of the FMS dispatching decision analysis 1

Buffer with pallet (YnLs=HaPal)	Pallet is waiting for moving (YnMovCmd=Yes)	Result	Exposition
T	T	T	There is a pallet on the buffer and is waiting for moving.
T	F	F	There is a pallet on the buffer, but will not move.
F	T	F	No pallet is on the buffer, but is waiting for moving. (Not feasible)
F	F	F	No pallet is on the buffer nor is waiting for moving.

Table 3

Illustrated examples of the FMS dispatching decision analysis 2

Part has next process (CurOpSeq<TotalOp)	Buffer of next processing station is empty (Next Buffer=Idle)	Result	Exposition
T	T	T	Part has next process and buffer of next processing station is empty. (Pallet will move to the next buffer)
T	F	F	Part has next process and buffer of next processing station is occupied. (Refer to Table 4)
F	T	F	Part has finished all processes and buffer of next processing station is empty. (Return to AS/RS)
F	F	F	Part has finished all processes and buffer of next processing station is occupied. (Return to AS/RS)

Table 4

Illustrated examples of the FMS dispatching decision analysis 3

Buffer of next processing station is occupied (Next Buffer<> Idle)	Pallet-Number equals to Buffer-Number (PalNo=CurBufferNo)	Result	Exposition
T	T	T	Do not move
T	F	F	Return to AS/RS
F	T	F	Move to next buffer
F	F	F	Move to next buffer

The integrity and suitability of this integrated IEMPD and FMS structure provides a manufacturing decision-making environment. Further, simulation can be implemented by production engineers to validate the performance of the assignment dispatching rules. Knowledge rules of the manufacturing process can be reused by different projects. Various scenarios analysis of the projects based on the similar integrated IEMPD and FMS structure can be conducted to improve current manufacturing operations and to set the appropriate response plans confronted by the future development. Therefore, knowledge-based information (production manufacturing know-how and experiences) can be shared, accumulated, represented, and represented. The decision makers, i.e., production engineers, can efficiently conduct and evaluate the system status for optimal decision making.

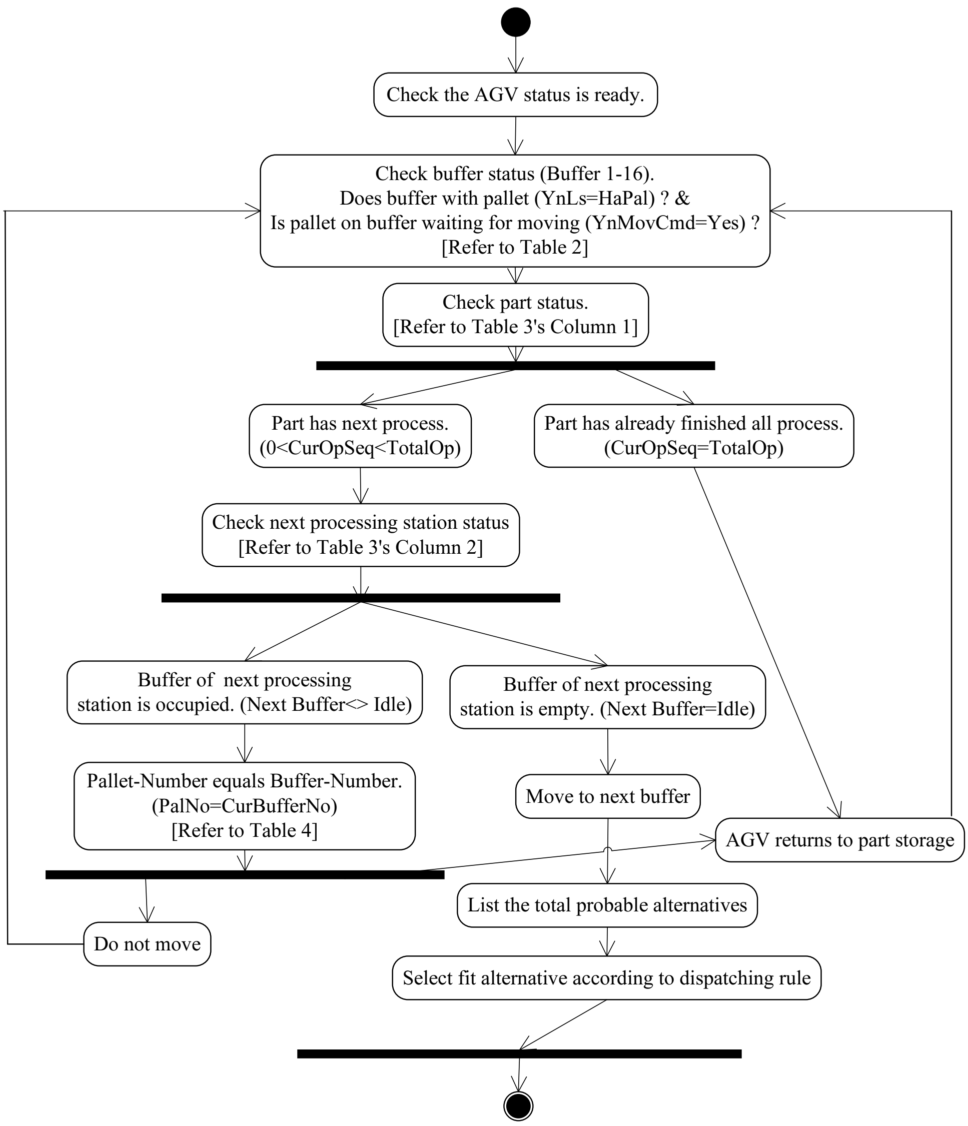


Figure 9. The dispatching inference process of the integrated IEMPD and FMS structure.

5 CONCLUSION

As flexible and agile manufacturing become prevalent, the ability to describe the short term future performance of a manufacturing system becomes critical. The prototype IEMPD system, which combines rule knowledge base and the inference engine, is developed and implemented. Knowledge engineers can construct, manage and maintain knowledge rules via the platform easily and flexibly to satisfy production or maintenance decision supports. Further, this research uses an example machine shop FMS to demonstrate the intelligent production supporting capability. And, the simulation methodology can be further implanted to compare the result of the dispatching performances. The prototype system enhances the agility of the manufacturing processes by supporting the decision making of equipment and production patching.

6 REFERENCES

[1] Balakrishnan K & Honavar V. (2003) Intelligent diagnosis systems. Artificial Intelligence Research Group, Department of Computer Research Group, Iowa State University, U.S.A..

[2] Bengtsson M. (2003) Standardization issues in condition based maintenance. Proceedings of the 16th International Congress on Condition Monitoring and Diagnostic Engineering Management, Vaxjo University, Sweden, August, 27-29.

[3] Byman B., Yarborough T, Carolefeld RSV & Gorp JV. (2000) Using Distributed Power Quality Monitoring for Better Electrical System Management. IEEE Transactions on Industry Applications, 36(5).

[4] Cabitza IF. (2006) Introduction to JESS: the Java expert system shell. Retrieved January 2, 2006, from http://www.mac.disco.unimib.it.

[5] Changchit C & Holsapple CW. (2001) Supporting managers' internal control evaluations: An expert system and experimental results. Decision Support Systems, 30(4), 437–449.

[6] Chen A, Guo RS & Yeh PC. (2000) An Effective SPC Approach to Monitoring Semiconductor Quality Data with Multiple Variation Source," Semiconductor Manufacturing Technology Workshop, 279-282.

[7] Chiang ATA, Trappey AJC & Ku CC. (2004) Using knowledge-based intelligent reasoning to support dynamic collaborative design. Proceedings, Asia-Pacific Industrial Engineering and Management Systems (APIEMS). December 12-15, Gold Coast, Queensland, Australia.

[8] CLIPS. (2005) Retrieved December 30, 2005, from http://www.ghg.net/clips/CLIPS.html

[9] Eriksson H. (2003) Using JessTab to integrate protégé and jess. IEEE Intelligent Systems, 18(2), 43–50.

[10] Jess. the Rule Engine for the Java Platform. (2006) Retrieved January 5, 2006, from http://www.jessrules.com/jess/docs/index.shtml

[11] Jovanovic J, Gasevic D & Devedzic V. (2004) A GUI for Jess. Expert Systems With Applications, 26(4), 625-637.

[12] Lamma E, Maestrami L, Mello P, Riguzzi F & Storari S. (2001) Rule-based Programming for Building Expert Systems: a Comparison in the Microbiological Data Validation and Surveillance Domain, Second International Workshop on Rule-Based Programming Affiliated with PLI 2001.September 4, Firenze, Italy.

[13] Lee J. (2001) A Framework for Web-Enabled E-Maintenance System. Wisconsin Distinguished and Rockwell Automation Professor & Director of NSFI / UCRC for Intelligent Maintenance System (IMS), University of Wisconsin-Milwaukee, USA.

[14] Limanond S, Si J & Tsakalis K. (1998) Monitoring and Control of Semiconductor Manufacturing Processes. IEEE Control Systems Magazine, 18, 46-58.

[15] Lin YT, Tseng SS & Tsai CF. (2003). Design and implementation of new object-oriented rule base management system, Expert Systems with Applications, 25, 369–385.

[16] Ngai EWT & Cheng TCE. (2001) A knowledge-based system for supporting performance measurement of AMT projects: A research agenda. International Journal of Operations and Production Management, 1(1–2), 223–234.

[17] Preece AD, Grossner C & Radhakrishnan T. (1996) Validating dynamic properties of rule-based systems. International Journal of human-computer studies, 44, 145-169.

[18] Rao MP, Miller DM & Lin B. (2005) PET: An expert system for productivity analysis. Expert Systems with Applications, 29, 300–309.

[19] Reichgelt H. (1991) Knowledge representation, an AI perspective. Norwood, NJ: Ablex Publishing Corporation.

[20] Roesner H. (1988) Expert systems for commercial use. Artificial Intelligence and Expert Systems, 35–59.

[21] Santhanam R & Elam J. (1998) A survey of knowledge-based systems research in decision sciences (1980–1995). Journal of the Operational Research Society, 49(5), 445–457.

[22] Shal SME & Morris AS. (2000) A Fuzzy Expert System for Fault Detection in Statistical Process Control of Industrial Processes. IEEE Transactions on Systems, Man, and Cybernetics, Part C, 30, 281-289.

[23] Souza MAF & Ferreira MAGV. (2002) Designing reusable rule-based architectures with design patterns. Expert Systems with Applications, 23(4), 395–403.

[24] Turban E & Aronson JE. (2001) Decision support systems and intelligent systems, 6th Ed. Upper Saddle River, NJ: Prentice Hall Inc.

[25] Wang H & Wang C. (1997) Intelligent Agents in the Nuclear Industry. IEEE Computer, 30(11), 28-31.

[26] Wang Z, Liu Y & Griffin PJ. (2000) Neural Net and Expert System Diagnose Transformer Faults. IEEE Computer Applications in Power, 13, 50-55.

[27] Wagner WP, Otto J & Chung QB. (2002) Knowledge acquisition for expert systems in accounting and financial problem domains. Knowledge-Based Systems, 15(8), 439–447.

[28] Wagner WP, Najdawi MK & Chung, QB. (2001) Selection of knowledge acquisition techniques based upon the problem domain characteristics of production and operations management expert systems. Expert Systems, 18(2), 76–87.

[29] Yao YH, Lin GYP & Trappey AJC. (2006) Using Knowledge-Based Intelligent Reasoning to Support Dynamic Equipment Diagnosis and Maintenance. International Journal of Enterprise Information Systems, 2(1), 17-29.

[30] Yao YH, Trappey AJC, Ku CC, Lin GYP, Tsai JP, Ho PS & Hung Benson. (2005) Develop an intelligent equipment maintenance system using cyber-enabled JESS technology. Proceedings, International Conference on Hands-on Intelligent Mechatronics and Automation (HIMA2005), IEEE, July 10-12, Taipei, Taiwan., 927-932.

Acknowledgments

This research is partially supported by the Industrial Technology Research Institute (ITRI) and the National Science Council (NSC), Taiwan. We also like to thank Dr. K.Y. Chen at the National Taipei University of Technology (NTUT) for providing us with the machine shop FMS case scenario.

ASSET MANAGEMENT KPIS, VALIDATED, ANALYSED AND PREDICTED UTILISING SELF LEARNING HYBRID MODELLING IN REAL TIME

Don Sands [a] and Glen MacElroy [b]

[a] *Technical Director, Synengco Pty Ltd, 82 Eagle Street, Brisbane, Queensland, Australia.*

[b] *CEO, Synengco Pty Ltd, 82 Eagle Street, Brisbane, Queensland, Australia.*

Abstract: Asset managers agree that real-time operational, economic and asset life Key Performance Indicators (KPIs) are essential for world class asset management. In seeking to achieve this, most enterprises, despite advances in information technology, still struggle with the challenge of obtaining valid (a true reflection of what is happening) and reliable (repeated measures having a low error factor) data. Enterprises have developed extensive monitoring and analysis systems that produce more data but have not significantly improved actionable information or reliable measures. As a result they are often unable to reconcile or understand the data, which is being presented at greater speed and in larger quantities. The barriers to achieving world class Asset Management KPIs include data invalidity, incompatible data structures, out of date models, analysis disconnected to the asset operation, information overload and delays in analysis of information. This paper outlines how the use of Hybrid Modelling, model based data validation and operations modelling has overcome these barriers to provide a comprehensive self learning, real-enough-time analysis, reporting and prediction system. The system provides valid and reliable operational and financial measures at all levels of the enterprise to be used for asset management, performance analysis, scenario planning, forecasting and prediction of events.

Key Words: Synengco, Hybrid Modelling, Asset Management, KPI.

1 CONTEXT

The objective of asset management is to maximise the outcome of the asset under management, which requires the optimisation of the cost of constructing, maintaining and operating the asset against the value of the output of the asset, as shown in Figure 1. To achieve optimal asset management disciplines and systems have been developed to help manage and optimise the costs of these individual components. For example Project Management for managing construction and asset modifications and Reliability Centred Maintenance (RCM), Computer Maintenance Management Systems (CMMS) etc for maintenance. Traditionally in referring to asset management emphasis is placed on maintenance systems and in the design element of construction while the systems associated with operations activities have a much lesser focus. This lack of focus by asset management on the operation of an asset occurs not due to the lack of recognition of the significant impact asset operations has on the asset outcome but more due to the lack of systems and frameworks in which analyse the impact of operations on the asset outcome and the other elements of asset management.

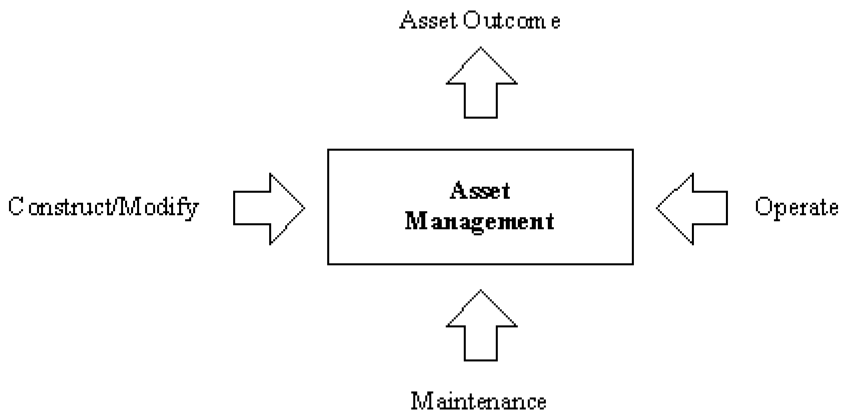


Figure 1: Asset Management Relationship

This paper will outline a methodology and requirements for the analysis of asset operation and how the asset operations is linked to, analysed and optimised with the other inputs into Asset Management. The paper is broken into three parts:

1. The characteristics of a good operations asset management system;
2. The way in which SentientSystem achieves the characteristics of a good operations asset management system; and
3. Case studies illustrating the application of SentientSystem to specific client issues.

2 CHARACTERISTICS OF A GOOD OPERATIONS ASSET MANAGEMENT SYSTEM

2.1 Asset Management

Asset management is the long-term management of an asset and generally covers the entire life cycle of the asset from concept to disposal. Effective asset management systems consider the following long and short term aspects of asset management in an integrated manner:

- ***Long term asset management issues*** include plant design, commissioning, plant degradation, staff turn over, replacement and refurbishment decisions.
- ***Short term asset management issues*** include short term production requirements, plant unavailability, critical incidents and planned and unplanned maintenance requirements.

A balance has to be struck between the long and short term aspects of asset management. This balance can be achieved through detailed knowledge of the asset and its workings over the short and long term. Support systems for asset management should use the available knowledge (see below) to provide a systemic long-term management system.

2.2 Asset Knowledge

Effective asset management uses knowledge (often captured over many years) of the plant and similar plant to decrease the life cycle cost, increase production, decrease risk and ensure the sustainability of the business. Over the life cycle of an asset, the following sources of knowledge, can be used for more effective asset management decisions:

- Plant design knowledge
- Commissioning knowledge
- Operational knowledge (long and short term)
- Maintenance effectiveness knowledge (of actual and similar plant)
- Overhauls/refurbishment knowledge
- Benchmarking knowledge
- Plant degradation knowledge (of actual and similar plant)
- Plant improvement knowledge (of actual and similar plant)
- Reliability knowledge (of actual and similar plant)
- Environmental impact knowledge (of actual and similar plant)

Effective asset management should, therefore, have a framework to manage the many and disparate sources of knowledge relating to the assets.

2.3 Precise Information

Many large cost decisions are required during the life of a large asset. These decisions are generally made based on information presented to a decision maker at the time the decision is required. Typically:

- ***maintenance and refurbishment*** decisions are based on how a plant is degrading and the impact maintenance and refurbishment activates have on the asset.
- ***abnormal operation and early warning for critical incidents*** decisions are generally related to how fast an asset is changing and the time to when this change will have secondary influences.
- ***optimisation*** decisions are based on how much the asset changes as a result of small changes to controllable inputs.

Precise information regarding changes (meaning small changes that are detected and repeatable) is required for all of the above decisions. As a result it is detecting and quantifying the change rather than the absolute value which is most import from a decision making process particularly from the operational asset management aspect.

It is also important that conflicting information be identified as early as possible. Unresolved conflicting information can have a major detrimental impact on decision making quality. Resolved conflicting information can often assist in the quality of the decision making process.

In summary information presented for use in effective asset management decisions must be precise with conflicting information clearly identified.

2.4 Abnormal Operation Detection

The abnormal operation of plant and processes is a leading cause of economic loss and critical incidents and an extremely important part of asset management. Industrial statistics estimate the economic impact due to abnormal operations and situations to be around $20 billion in the petrochemical industry alone (Nimmo, 1995). Abnormal operation analysis is traditionally associated with large complex processing plants (such as in the petrochemical and power generation industries).

There are considerable technical challenges in developing the systems necessary to achieve the early detection of abnormal situations and the identification of the root causes of these situations including:

- the complexity of process dynamics
- lack of adequate models
- incomplete and uncertain data
- diverse sources of knowledge
- amount of effort and expertise required to develop and maintain the systems
- the large operating ranges that the processes are subject to under normal operating conditions

Any mechanism that detects abnormal operations, predicts failure and provides root cause analysis for the detected or predicted events is a useful tool for improving asset performance.

2.5 Continuous and Discrete Processes

Asset management involves both continuous and discrete processes. Product production generally involves a continuous or semi continuous process while maintenance and refurbishment processes are generally a discrete process. Effective asset management therefore requires a mechanism to allow the integration of both continuous and discrete processes within the same decision framework.

2.6 Modelling

Modelling is an effective mechanism to capture knowledge. A great deal of historical scientific research is captured in fundamental and empirical models. Models provide a mechanism for providing predicted outputs based on given inputs allowing cost effective deployment of knowledge that may have taken many years of research to discover and test. Modelling also allows complex processes to be broken down into smaller more easily understood components.

The two basic forms of modelling, fundamental and empirical based models, and their advantages and disadvantages are discussed below.

2.6.1 Fundamental Models

Fundamental models are based on processes. In its simplest form a fundamental model could be a simple equation that represents a fundamental relationship like Power = Current x Voltage.

More complex fundamental models include thermodynamic models, network models, chemical reaction models etc. These rely on the mathematical relations that exist between variables and a consistent model tries to express these relations in a closed form. Model are developed using first principles – heat, mass, momentum balance, reaction kinetics etc. or input-output data and can be dynamic, static, linear or non-linear.

The major advantages of using fundamental models are that the system only requires small historical data sets to allow operation and deviations can be deterministically analysed.

The implementation of these systems is, however, problematic due to the difficulty of obtaining accurate and stable process models, the susceptibility of the outputs to the drift and failure of the inputs, the sudden process reconfiguration under normal operating conditions as well as the expected degradation and renewal of the process through maintenance activities. Further the models may not account adequately for the dynamic performance of the system and therefore produce a large number of fault detections (false positives) which are, in effect, normal dynamic behavior. Another disadvantage of using fundamental model systems are that they generate large amounts of data and hypotheses, which require large amounts of time and expertise to analyze. As the models are based on process fundamentals they also tend to be unique solutions for that process and resulting in high cost and little transportability.

Fundamental models are a powerful tool for stable slow changing systems and can provide excellent early detection, diagnosis and fault discrimination. They are, however, susceptible to process changes, multiple faults and may not prove very robust in an industrial environment where the input data is often noisy and unreliable.

2.6.2 Empirical Models

Empirical models are based on data. In its simplest form an empirical model could be a simple curve. More complex empirical models include neural networks, Bayesian networks etc.

The primary advantages of empirical models stem from the fact that little or no knowledge of the fundamental principles of the process are required to apply the techniques. These models can, therefore, be applied to any process and at a reasonable cost. The system outputs are generally readily understood and interpreted, at least at the superficial level, making the system readily actionable by a range of personnel.

The disadvantages of the system are that they require large data sets both in process points as well as history. These data sets provide a technical challenge to manage and process in real time. The training data (base line) may result in the system masking faults as the data may include the fault or such a large range of operations that the fault is indiscernible with the normal operation of the plant. A further disadvantage of these systems is that they do not inherently provide a diagnostic methodology to identify the root cause of problems.

3 SENTIENTSYSTEM

With the characteristics of an effective asset management system in mind, Synengco developed SentientSystem. SentientSystem is a system that brings the key aspects of effective asset management into an integrated framework. It is designed to be used over the whole of life for an asset from initial concept, commissioning, operation, refurbishment through to disposal. It can be used for design, performance monitoring, abnormal event detection, root cause analysis, risk management, knowledge management and short and long term production costing. It can be used on a single plant or to integrate a portfolio of plant. Its focus is on precision and reconciliation of information used for operational asset management decisions.

The SentientSystem framework allows for traditionally disparate data and analysis systems to work together in a synergetic manner. As more data, information and knowledge is available it can be applied in a consistent manner, reconciling the new information and knowledge with existing information and knowledge. The system provides a consistent mechanism across the data for extracting features from a massive amount of collected and generated data, detection of change from variable “normal” operation and a consistent mechanism for alerting personnel when a notifiable change is detected.

3.1 Knowledge Management

A large amount of knowledge is generated during the life of a large asset. During the design process many fundamental models are developed and optimised to deliver the most cost effective solution for the predicted future environment. During construction knowledge is generated on construction costs, plant changes, shortcuts taken and designs that could not be implemented. During commissioning there is traditionally a large amount of parametric testing conducted. Operations generate knowledge of the market, plant issues and raw materials issues. Disposal generates knowledge of asset worth and disposal costs. Traditionally much of this knowledge is captured but is not readily available to other parts of the life cycle. SentientSystem provides a mechanism to capture this knowledge and also similar knowledge from other plant, benchmarking studies and research and deploy it in real time.

SentientSystem captures knowledge generated from construction, commissioning and operations and checks whether there is a conflict with existing knowledge.

3.2 Hybrid Model

We found from experience that fundamental and/or empirical models were not effective on their own for asset management. Fundamental models are excellent for design purposes but are difficult to maintain operationally and do not cater for normal plant degradation over time. Empirical models lose a lot of available knowledge and miss important characteristics. They do, however, allow for self learning and are much easier to maintain.

As a result, the basis of SentientSystem became a hybrid modelling framework using a mix of fundamental and empirical modelling techniques that can handle both continuous and discrete processes in a systemic and auditable framework. This allows knowledge to be associated with the model in many different ways, from fundamental and empirical knowledge sources to rules of thumb and captured prediscovered cases. The model allows for self-learning of fundamental models characteristics using empirical techniques but still utilises fundamental knowledge to provide rapid learning and the ability to check the validity of the data being presented.

SentientSystem’s hybrid modelling technique was also designed for robustness and precision meaning that the information is reconciled as part of the solution using residuals. Residual based techniques are used within the fundamental models to improve precision and robustness of solutions by ensuring that the models are manipulated to achieve a solution that provides the lowest practical residual within the fundamental model to get the closest solution that represents the data presented. This technique not only ensures a highly repeatable result but also increases the early detection of faults through monitoring residual amounts for each solution. The residual techniques can also provide strong diagnostic capabilities in analysing a failure as it assists in identification of the component elements that required manipulation to obtain a solution. Online reconciliation has increased the precision of a solution, ensures the solution is feasible from a fundamental aspect and that small changes to the process are able to be determined in a repeatable and accurate manner despite noise on the input signals. This allows small and subtle changes to be detected early.

When developing the fundamental model, a review is conducted of the process to identify where accurate fundamental models can be applied for components within the process. These components are then built. Where fundamental models are not available or are unreliable, statistical models are built in conjunction with fundamental models for these components. Both the fundamental and statistical component models are then integrated to complete a total process model.

Many of the challenges of the application of fundamental models to industrial and commercial applications lies not within the identification of the appropriate mathematical model, but the identification of the system parameters for this model. The SentientSystem fundamental model uses a self-learning framework, which matches the mathematical model to the process data and self learns the system parameters. These self-learnt parameters are then validated within a process knowledge set to prevent solutions being realised that are not congruent with the process.

This self-learning framework continues to operate throughout the life of the system providing an adaptation mechanism such that the model accurately reflects the process as it changes with time and use. Analysis of the changing system parameters is itself a highly useful diagnostic tool.

The fundamental model also provides the structure and data hierarchy to improve the ease with which process and components can be analysed.

3.3 Virtual Instruments

Virtual instruments are a derived piece of information that is not available in a raw form but is closer to the decision making process. A common example of a virtual instrument is the fuel efficiency figure for modern motor vehicle where a fuel consumption (l/hour) and speed (km/hr) is translated to a l/100 km which is closer to a decision of should I increase or decrease speed to improve efficiency. A virtual instrument may also be a measurement within processes that cannot or does not have any instrumentation. However, by using other instruments within the same process and a model of the process, the virtual instrument value can be derived. An example could be a virtual instrument of the mechanical power output from an electric motor using electrical measures on the power supplied and a model of the electric motor.

SentientSystem provides a mechanism to allow virtual instruments to be added to any part of the process. The virtual instruments have a model (empirical, fundamental or hybrid as appropriate) associated with them. There can also be multiple virtual instruments generated from a single model (multiple inputs and multiple outputs). Virtual instruments can also be used for prediction.

3.4 Data Validation

Data validation is possibly one of the largest reason for actions not occurring as a result of diagnosis and analysis of real-time data sources. Generally there is a large amount of real-time data upon which to make the decision. The data is, however, sometimes incorrect due to instrument errors, incorrect data entry and/or deliberately misleading information. The SentientSystem framework allows data to be validated as soon as it comes into the system through many different mechanisms. One of the more powerful data validation mechanisms is the ability to check the relationship of the data with other sources of data associated with the same process using virtual instruments. This provides for real-time reconciliation and data validation which is extremely important in any asset management system.

3.5 Key Performance Indicators

Key performance indicators are basically virtual instruments on the key measures that indicate the health of the business process. Sometimes there are direct measures but more often the key performance indicators are an aggregation of a number of measures and algorithms and provide a mechanism to compress lots of information into a much small number of KPIs. Traditionally, due to the data compression mechanism, a change in a KPI is difficult to root cause without further analysis. SentientSystem provides for root cause analysis as indicated below for deviations in KPI.

3.6 Monitoring, Abnormal Event Detection And Root Cause Analysis

Within SentientSystem there is a vast amount of data available for analysis. It is typically in a hierarchical plant model and includes all the current raw data, the virtual instrumented data and historical data (sometimes over many years). Monitoring and abnormal event detection is achieved in multiple ways. The first is to generate virtual instruments (expected behaviour) for the process parameter being monitored and compare it with the actual instrumented value. If there is a significant deviation between the virtual and actual instrumented value an alert can be sent out. The other methodologies are based on a comparison of the current data signatures with historical data signatures. If the current data signatures deviate significantly from the baseline signatures, then an alert can be sent out. Root cause analysis can be achieved by comparing the current data signature with the signatures of known problems. There are many techniques available for developing signatures which are based on advanced analysis, classification, machine learning, statistical and data mining technologies.

Each of the following techniques can be implemented, depending on the evaluation of the particular process and detection and diagnosis required.

Principle Component Analysis (PCA): PCA carries out a statistical analysis on 'n' variables over a given time period, at a given interval, and transforms the 'n' process variables to be represented by 'm' principal components (which are dimensionless components that describe the variation between the relationship of process variables). There is a statistical link between the principal components and the process variable but this is not readily observed in the analysis of only the principal components.

The power of the PCA is that for a sensible set of data (data derived from an underlying process), the majority of the variance (>85%) is displayed in the first and second principal components. It allows you to study the variance of a process containing hundreds of variables by observing the changes to the first two principal components from a PCA study. This technique has been refined further by calculating the ellipses that contain 95% and 99% of the points and then repeating the analysis on a test set. By observing the instances that fall outside the target ellipses we can detect instances in time when the process was no longer operating as per the base case. This results in a technology that can easily detect abnormal operation.

PCA studies have more recently been utilised to also determine root causes, as the pattern of the PCA plot is characteristic of the root cause giving rise to the abnormal event. Development in this area is the focus of many organisations using PCA.

While the detection of events using the PCA is useful, additional techniques were developed to compliment the PCA to assist in root cause analysis of the out of control events. These techniques include Hotellings T^2 (TA^2), Partial Least Squares (PLS) and Squared Prediction Error (SPE). The TA^2 analysis provides a time series analysis of all the elements of deviation to allow a time analysis, the PLS is used to assess the significance of the deviation and the SPE provides a check that the analysis of the first 2 principal components is valid.

Slow Feature Analysis (SFA): This analysis is applied to remove the impacts of fast transients which can mask the slowly moving underlying drivers of a system. This technology was originally developed for the analysis of biological systems to determine the drivers for population increases and decreases and has been applied by SentientSystem to industrial processes to identify slow moving changes and provide early warning to these slower changes. This technology enhances abnormal event detection by removing the transients that provide false positives as well as providing early warning of process and plant degradation.

Independent Component Analysis (ICA): ICA looks for the statistical independence of the data and identifies the independent components that represent the raw data. This technique has been used extensively for blind source separation, where signals from various sources can only be measures as a mixture of the signals (an example is multiple conversations occurring at once with the microphones only able to measure the signals mixed together). ICA is used to separate the signals into their independent components. This means that the value of any one of the independent components gives no information to the value of another of the independent components (one conversation gives no information to the other conversation). By monitoring the independent components we can effectively monitor the source of the signal (one conversation), not the measurement of the signal (the microphone), and detect any changes that occur at the source.

Within an industrial process there are many sources of disturbances impacting on the measurement of the process such as ambient conditions, incoming material quality and plant degradation. The monitoring of the process may be through process variables such as pressures, temperatures, mass flows, etc. Through the application of ICA, it is possible to monitor the source disturbance instead of the raw measurement, account for known disturbances and monitor unaccounted disturbances. The monitoring of the unaccounted disturbances provides a mechanism for abnormal event detection.

Neural Networks (NN): This advanced multidimensional curve fitting technology is useful for providing a statistical representation of processes, which are difficult or impossible to create a fundamental model given the state of the model knowledge or the availability of data. The system statistically fits a multidimensional curve to the input and output data of the component under analysis. The curve is then available for simulation or predictive analysis where actual results are monitored against the prediction and deviations monitored and alerted to detect changes in process performance.

Partial Least Square Prediction (PLS): This advanced multidimensional analysis technique provides for the development of predictive models based on statistical relationships learned from historical data. The prediction of measurements compared to the actual measurements provides a mechanism to detect changes in the process and a method of abnormal event detection.

Classification Models: Classification is a method of data reduction and is used to classify information into distinct clusters that have certain characteristics. SentientSystem includes many classification techniques that can be applied to different problems. Use of classification is useful for abnormal event detection within a dynamic process through development of known operation classifications. This builds up classifications for known operations and anything that cannot be classified into known operation becomes unknown operations and the basis of abnormal event detection.

Time Series Analysis (TSA): TSA involves the analysis of data within a time dimension and is focused on trend and seasonal analysis. Seasonal analysis is the repeated cyclic information that has many different time periods from seconds to months or even years. Trend analysis is the long-term changes that do not have a cyclical nature. By identifying the individual trend and seasonal components, the underlying and independent phenomenon can be extracted, predictions of the time series variables can be achieved and abnormal events can be detected as a result of changes to the seasonal and trend components or as a deviation from the predicted and actual results. SentientSystem utilises a range of techniques for TSA.

3.7 Simulation and Optimisation

Simulation of the models is a powerful asset management tool. Through the manipulation of the input variables and model parameters (such that the predicted outputs match the actual outputs), a rapid diagnosis of possible fault states can be obtained. Simulation allows for optimisation by manipulating controllable parameters such as maintenance, component replacement and day to day operations, such that a more desirable outcome is achieved. Simulation can also be used to simulate plant failures and develop plant fault signatures for root cause analysis without having to experience the fault first hand. The SentientSystem simulator utilises the same fundamental and costing models that are used to monitor the process ensuring that the predicted values from the simulator are based on the self learnt system parameters reflecting the current process reality.

3.8 Costing

In prioritising and assessing fault detection, operating regimes, asset degradation for further analysis and action, the costing model provides key information that is actionable by all users. The costing model is integrated with the fundamental model such that an integrated financial and process solution is obtained. This ensures that any change to the asset has an integrated engineering/physical and financial analysis. This provides for the system performance to be monitored via either domain, financial or process, and allows for resource allocation and prioritisation to be made across the organisation.

The costing analysis can be componentised into a range of costs including conversion cost, capacity, asset consumption, maintenance, service quality, environmental and regulatory costs.

3.9 Outcomes and Benefits

The benefits for Asset Management from using a tool such as Sentient System are:

- ***Data Validation:*** With effective data validation decisions can made quickly and effectively without having to worry about if the data is correct.
- ***Early Warning of Problems:*** This prevents collateral damage from plant failures and reduces the cost of failure. It also allows planning to take place before a failure, thus reducing the cost and time to repair.
- ***Key Performance Indicators:*** Key performance indicators provide a means of measuring the overall performance of the asset and allow the monitoring of the effectiveness of asset management strategies.
- ***Optimisation:*** Optimisation results in reduced asset consumption and maintenance requirements while improving production.

4 APPENDIX A: CASE STUDY SENTIENTSYSTEM OPTIMISATION

The following shows how the SentientSystem™ Optimisation can be used to improve asset life and reduce failures. A plant in the US was experiencing plant failures from to the size and frequency of thermal excursions on critical metal components due to cleaning activities. This thermal cycling was contributing to metal fatigue and failures. SentientSystem was used to monitor, optimise and control the cleaning activities.

A number of different technologies from SentientSystem were used to address this problem. Due to the arduous nature of the environment, instrumentation to measure the thermal cycles would be very expensive, unreliable and would have required a unit outage to achieve. SentientSystem, instead used existing instrumentation, its precision model and virtual instrument to monitor the thermal cycles.

Once the thermal cycles were monitored, the self learning aspect of SentientSystem was used to characterise the impact that the cleaning action (discrete process) had on the metal temperatures (continuous process). This influence changed with different feed material (semi-continuous process) and different production levels (continuous process). As a result the self-learning aspect had to cater for these "normal" changes within the process and separate out the cleaning influence from all the other disturbances.

With the cleaning characteristic self-learned, SentientSystem optimisation was used to optimise the cleaning regime with the primary objective of decreasing asset consumption and improving reliability. A secondary objective was to maintain or improve the efficiency of the plant. To provide the balance between potentially conflicting objective functions SentientSystem costing was used to weight the influences according to their relative value/cost to the plant. The optimisation then only had to minimise the cost of the solution to achieve the most balanced outcome for the plant.

The outcome for the plant was an improvement of 40% in mean time to failure for the plant, an improvement in the efficiency of the plant and a reduction in the cleaning resources required by 50%.

As a result of using SentientSystem different technologies, this opportunity was realised without additional instrumentation or a plant outage. The resultant low cost and impact of installation meant the project had a true payback period of less then 6 months.

5 APPENDIX B: CASE STUDY SENTIENTSYSTEM EARLY WARNING

The following shows how the SentientSystem™ MSPC can detect a tube leak in a boiler by monitoring the thermal conductances as calculated by SentientSystem™ Process Analysis. The first detection of the leak occurred 4 days and 12 hours prior to the unit coming off line and was over 2 days earlier than the confirmation through the traditional fault detection system.

The timed PCA in Figures A1 to A4 graphically shows how the leak initiated and then propagated. Once detected, Figure A5 shows how the Variable Contribution Tool identified, which process variables are contributing to the abnormal operation. Through knowledge of the boiler layout and that the SSH, PSH and Economiser are all contributing to the variance we can conclude that the element prior to these has the leak as additional energy is being placed into the gas stream which has increased the thermal conductances of the SSH, PSH and Economiser.

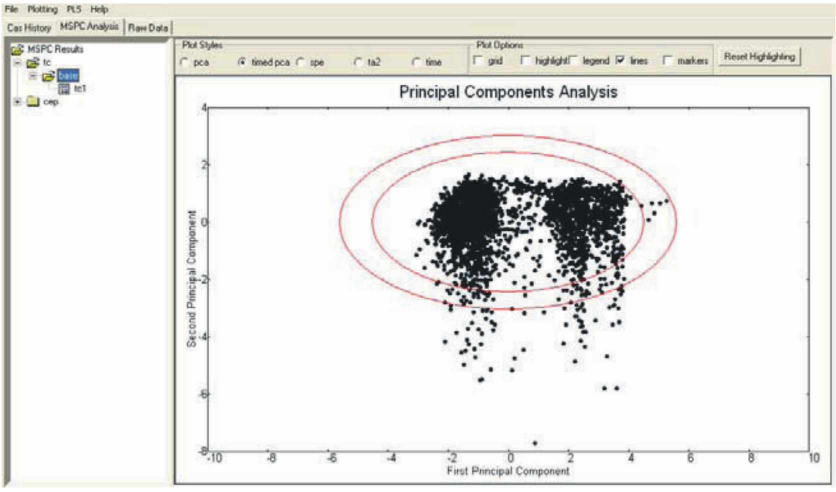


Figure A1: Base case using thermal conductances of boiler

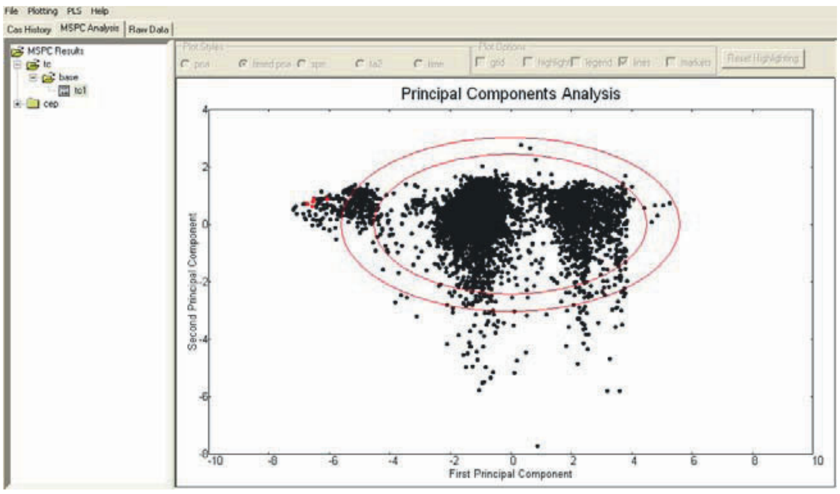


Figure A2: Test case showing start

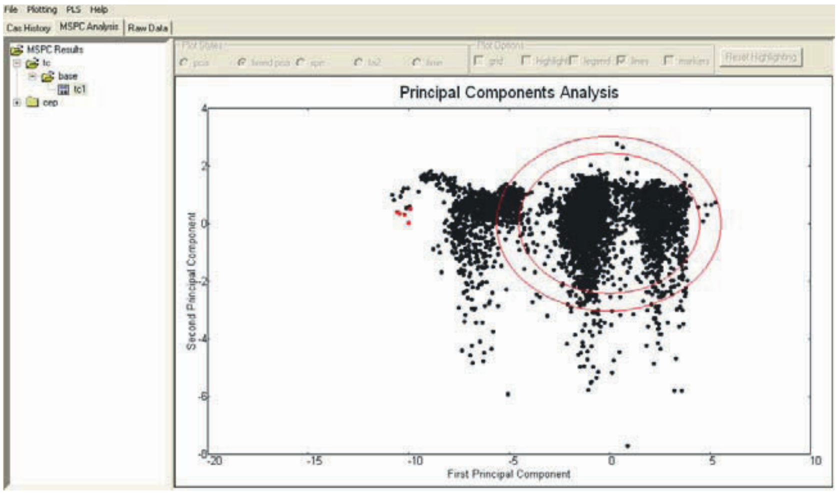


Figure A3: Test case showing start tube leak progression

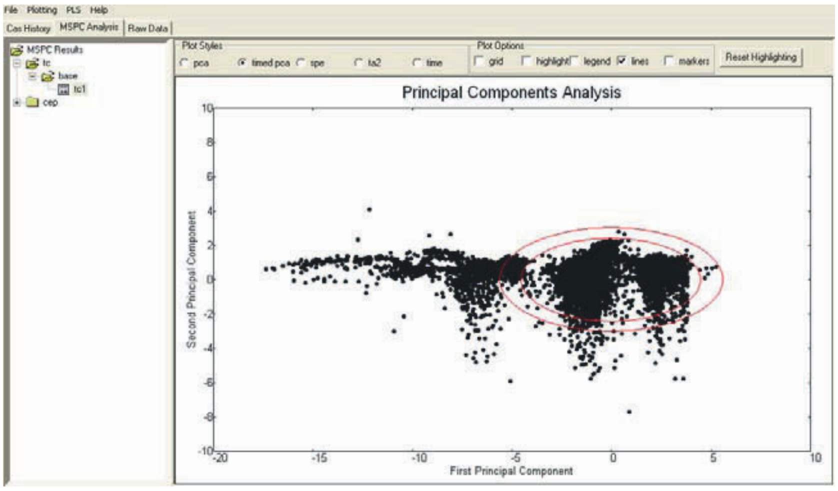


Figure A4: Test case showing final stages

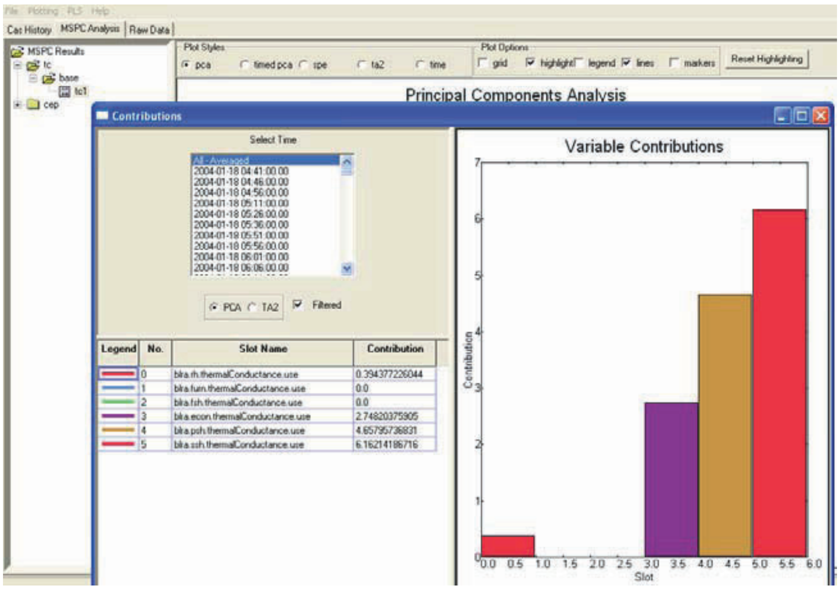


Figure A5: Determining the Process Variables Contribution to the Deviation

6 REFERENCES

[1] Venkatasubramanian, V. and Dash, S. *"Challenges In the Industrial Applications of Fault Diagnostic Systems"*, Laboratory for Intelligent Process Systems, School of Chemical Engineering, Purdue University, W. Lafayette, IN 47907, USA.

[2] Nimmo, I *"Adequately address abnormal situation operations"*, Chem. Eng. Prog. 91(9), 36-45, 1995.

[3] Lapp, S. A. and Powers, G. A. *"Computer aided synthesis of fault trees"*, IEEE Trans. Reliability, 37, 2-13, 1997.

[4] Iri, M., Aoki, K., O'Shima, E. and Matsuyama, H. *"An algorithm for diagnosis of system failures in chemical processes"*, Compt. Chem. Engng., 3, 489-493, 1979.

[5] MacGregor J. F., Christiana, J., Costas, K. and Kotoudi, M., *"Process Monitoring and diagnosis by multiblock PLS methods"*, AIChE J, 40(5), 826-838, 1994.

[6] Kresta, J. V., MacGregor J. F. and Marlin, T. E., *"Multivariate Statistical Monitoring of Processes"*, Can. J. Chem. Eng., 69, 35, 1991.

ORGANIZATIONAL MANAGEMENT OF PROJECT AND TECHNICAL KNOWLEDGE OVER FLEET LIFECYCLES

William P. Hall [a], Garry Richards [b], Carl Sarelius [c], and Bill Kilpatrick [d]

[a] *Tenix Defence, Williamstown, Vic. 3016, Australia.*

[b] *Tenix Defence Marine Division, Williamstown, Vic. 3016, Australia.*

[c] *Tenix Defence Land Division, Wingfield, S.A., 5013, Australia*

[d] *Tenix Defence Marine Division, Williamstown, Vic. 3016, Australia*

Abstract: Complex engineered products are all knowledge intensive. It is crucial for suppliers and operators of such "fleets" of equipment to manage and minimize product costs and risks. Many schedule delays, cost overruns, accidents, excessive operating costs, and premature product failures result from ineffective management of product data, information, and knowledge. This paper reviews solutions based on integrating structured authoring and product lifecycle management systems and data warehousing implemented by a large project engineering and management organization and the development of an alliance organizational form to support the product to reduce costs, risks and hazards through comprehensive and coherent management of project data, information and knowledge.

Key Words: Life cycle knowledge management, complex products, documentation, risk minimization

1 INTRODUCTION

Complex, long-lived products are failure prone and sources of costs and risks for organizations that procure and operate such things or build, supply and support them. Some problems relate to poor management of personal knowledge (Nousala et al. 2005). However, the points of failure leading to visible cost and schedule blowouts, in-service failures or accidents often derive from errors appearing first in project documentation and data, possibly as far back as client requirement development and contract negotiation. One key to minimize unanticipated costs and problems in service is to provide coherent visibility and control of the development of project requirements and their elaboration into documentation for engineering, production and maintenance activities. Another is to integrate lifecycle logistic support considerations into the systems and design engineering processes. A third is to provide engineering configuration control over all aspects of proposed changes to ensure that impacts on associated data and documentation deliverables are carefully considered before any change is allowed to proceed. A final key is to ensure rapid feedback and analysis of problems arising in production and in-service operations. This paper describes how Tenix Defence integrated a range of software into an increasingly comprehensive knowledge management infrastructure to satisfy requirements of production, operating and support organizations to better manage engineered fleet products.

1.1 Project life cycles — What can go wrong?

Almost anything relating to a complex product can go wrong, sometimes disastrously so. Because faults in early documentation development, design and logistics analysis phases "only involve paper", such "latent errors" (Hopkins 2005) can escape notice or be hidden until products are fabricated and put into service. However, the cost and delay to correct mistakes on paper is insignificant by comparison to costs to correct mistakes after they have been fabricated in steel, so the benefits from identifying and remedying such mistakes before they go into production should be high. The construction/ production phase of a large project may span a decade or more. In-service support may span several decades. Examples are ships, commercial and military aircraft, heavy vehicles, infrastructure plant and equipment, where units may be in service for several decades. Knowledge concerning the product must be actively maintained through this lifecycle for periods longer than the normal membership of a person in any the organizations.

By comparison to project successes, documented cases of mistakes and failures can be difficult to find. Many mistakes are covered up, rather than cherished and studied as learning situations (Argyris 1999; Ford & Sterman 2003). Yet, "Those who do not learn from history are doomed to repeat it" (attributed to George Santayana, 1863-1952).

Seeking to improve the performance of large Australian Defence projects with regard to cost and schedule blowout, difficulties with large Australian Defence projects are well documented. (e.g., McNally and Bates 1999; McNally et al 2005;

Brown 2002). One example of a project that has encountered problems throughout its lifecycle is the $A 5 billion Royal Australian Navy Collins Class submarine project to build six diesel powered attack submarines, documented in a report from the Australian National Audit Office (McNally 1997), the McIntosh/Prescott Report (1999) and a Parliamentary Research Paper (Woolner 2001). Many of the problems with the submarines traced back to the contract specifications - where the supplier explicitly followed the contract to deliver against specifications that were inadequate when the products were put into service. Also, monitoring procedures did not adequately highlight the issues prior to delivery of the subs.

Other issues tracing back to problems with knowledge work early in a project may become apparent only as operational failures. If appropriate knowledge doesn't exist, is incorrect, or isn't accessed and used when needed in engineering, producing or operating the product, huge costs can ensue and people may even die. Many equipment failures are resolved with little more than damage to the equipment itself. However, it is the nature of large and complex systems that some failures propagate catastrophically (Bea 2003). Such disasters often begin with engineering and design faults, or faults in maintenance and support. A "minor" equipment failure or operator error may be exacerbated by operators who lack immediate personal knowledge and cannot find explicit knowledge as to remedies that might bring a cascading sequence of failures under control. As failure becomes more extreme the chances that operators will have the necessary knowledge to stem the cascade declines.

Documented examples of these kinds of failures include: Three Mile Island (Kemeny 1979), Longford Gas Plant (Dawson and Brooks 1999), HMAS Westralia. (Yates 2000; Hope 2003; Royal Australian Navy 2004)., Petrobras's P-36 offshore semisubmersible oil production platform. (Rios 2002; Bea 2003; Dias and Valerio 2002), and Chernobyl (Cohen 1990).

A consideration in all these cases is that plant and equipment operators as individuals or teams can only ever be expected to have the knowledge required for their day-to-day activities and reasonably common occurrences in their heads (Hall 2003). And even then, to train the operators to that state of competency to deal with failures outside the normal operating envelope requires production of explicit training activities, technical manuals, simulators and a variety of other explicit knowledge-based materials to guide the training. Thus, beyond training for their common activities, to deal with infrequent, unusual and extreme events, plant and equipment operators require ready access to a much broader knowledge base than can possibly be held in personal memories — especially kinds of events that have the potential to become catastrophes if not dealt with correctly. Appropriate and correct documentation in a rapidly accessible (e.g., semantically indexed, retrievable) format can be a lifesaver — or at least essential to keeping the product in service for an economically long life. On the other hand if the documentation is available, but incorrect, equipment be prematurely destroyed or people die as a result.

1.2 Identifying and remediating project failures close to their sources

Appropriate infrastructure systems for managing and improving project technical knowledge via iterative feedback loops can reduce project risk (Manderson 2000). To try to stop engineering related faults before they reach production or affect in-service operations, the Australian Department of Defence has established a Technical Regulatory Framework (TRF) for major Defence materiel projects (Commonwealth of Australia 2004a) establishing management framework suitable for all kinds of major engineering projects, not just those in Defence. This accepts that engineering decisions involve risks, and establishes a principles-based framework to control and mitigate these risks during design, construction and maintenance. The TRF establishes two kinds of "authorised" agencies, the Authorised Engineering Organisation (AEO) and the Authorised Maintenance Organisation (AMO). Defence organizations such as System Program Offices (generally responsible for a particular fleet) and contractors (whether engineering or maintenance) are all required to obtain and maintain appropriate AEO or AMO certifications for their fleet-related design, engineering or maintenance activities.

An accredited organization is required to comply in four areas (Commonwealth of Australia 2004a), which just happens to summarize the four principle areas of knowledge management as defined in the Australian Standard (AS 5073(Int)-2003): *People*: Individuals within the organisation are qualified, authorized and have demonstrated competence to perform their designated activities. *Processes*: The organisation has, and uses, documented, controlled and approved plans, procedures and processes that conform to the TRF. *Systems*: The organisation has established and maintains applicable management systems appropriate to the type of work being performed. *Data or content*: The organisation uses relevant and authorized data appropriate to the activities being performed. Following Mensforth (2004), Navy's TRF for the engineering organization includes eight major elements:

- Managing risks
- Managing engineering change
- Reporting and investigating defective and unsatisfactory materiel
- Implementing maintenance management systems
- Proper delegations for engineering authority (to ensure that only appropriately qualified people review and authorize engineering changes)
- Proper authorization of engineering organizations (to ensure that it has accredited management systems, appropriately qualified and competent staff, and appropriately qualified management to enforce the integrity of all data)

- Implementation of technical data management system (data integrity, access & IP controls following defense standards, interfaces with other technical processes, procedures and management, interfaces to quality management system)
- Continuous enforcement of compliance with the TRF.

Physically transporting and tracking large volumes paper through the several layers of review approval can be major sources of cost and delay that has inhibited implementing strict TRF type controls. Systems able to electronically manage elements of content within structured documents, to distribute them at light speed and prompt for overdue evaluations or approvals greatly facilitate the implementation of a TRF.

1.3 Meeting product Requirements of the operating/client organization

The fundamental requirement of the operating organization with regard to the engineered product(s) is to have the product capabilities available and functional when needed. To meet this requirement, knowledge must be developed and provided to ensure that products:

- do what they are supposed to do reliably.
- are available for service when needed.
- are readily maintainable.
- are supportable.
- are operable within limits of human knowledge & capacity considering health, safety and operational knowledge issues.

Secondary considerations for the operator are to minimize product costs. For long-lived products such as ships and armored vehicles, support and maintenance costs for the product's life are typically several times the original acquisition costs. Thus, giving due regard to meeting the required capabilities, the product should be designed not only to minimize acquisition costs, but to minimize documentation, support & maintenance costs.

Adequate performance on all the above issues depends on the quality of authoring, management and transfer of technical knowledge from supplier to operators and maintainers. For the fleet operators, this technical knowledge must be:

- Correct in the sense that all knowledge deliverables have been subjected to stringent review and quality control to ensure their technical and operational accuracy in terms of the product's engineering and functional details.
- Consistent across the fleet to facilitate human training and computer systems interoperability
- "Applicable" to the configuration of the individual ship/vehicle
- "Effective" for the point in time re implementation of engineering changes, etc.
- Available to who needs it, when and where it is needed, and
- Useable, both in the sense that it is readily understandable by human operators and maintainers, and in the sense that it can be readily managed and processed in computer systems

To help minimize product costs, besides meeting the operators' requirements expressed above, the supplier has additional goals relating to assemble and produce the technical knowledge package for the product — bearing in mind the contradictions; e.g., to minimize production time (faster), deliver high quality (better), and to complete the work for a low cost (cheaper).

2 TENIX'S ARCHITECTURE FOR MANAGING FLEET DATA, INFORMATION, AND KNOWLEDGE

Tenix Defence is a large defense engineering and project management company that builds and/or maintains fleet "classes" of products or "platforms". The ancestral organization (now the Marine Division) was formed in the late 1970's to bid on the ANZAC Ship Project to build 10 frigates for the Australian and New Zealand Navies. Having won the ANZAC project, Tenix has grown and diversified into an organization of several divisions addressing the whole gamut of Australian Defence requirements, and a wider Group of associated companies covering a wide range from infrastructure engineering and project management to the development and commercialization of high technology and software products. Tenix's divisions now operate across Australia and in several countries overseas. Company success is likely due in part to a proven ability to deliver comparatively unproblematic products on cost and on schedule. At least some of this ability can be attributed to our increasingly sophisticated engineering knowledge management infrastructure.

The ANZAC Ship contract to build 8 frigates for the RAN plus 2 for New Zealand (ships 2 and 4 in sequence) was signed in 1989. The first ship was delivered in 1996. Further ships were delivered at the rate of about one per year, with the last delivered in 2006. The designed lifespan of a single ship is approximately 27 years, requiring ongoing and periodic maintenance support. Given changing strategic environments and rapid evolution of weapons and electronics technologies, major engineering changes will continue throughout each ship's life in service. Although built to similar plans, due to requirements of different clients, lessons learned, supplier changes, changing needs and technologies, and in-service engineering changes, no two ships are identical when delivered or remain static after delivery.

Besides responsibilities to build the ships, Marine was also contractually responsible to provide a comprehensive and highly knowledge intensive logistics support package. Unique aspects of the contract for this particular project are that:

- A stringent "fixed price" was negotiated for all of its aspects including the logistic support package.
- After 10 ship-years of in-service operations (Test, Evaluation and Validation period - TE&V) Marine had to prove to the client's satisfaction that the support package enabled the ships to meet contractual availability requirements for the combat system as a whole and for a number of individual "critical systems". For any shortfalls, Marine had to develop and implement acceptable remediation plans — all within the fixed price.
- A major milestone for the client's acceptance of Ship 5 was acceptance of the complete logistic support package, as summarized above. Failure to meet this milestone would trigger payment of major liquidated damages.

These unique aspects required Marine to develop innovative lifecycle knowledge management solutions that have now been extended and refined across additional projects. These solutions are based on a systems architecture for managing the gamut of information relating to fleet products to address in-house, client and operator needs. Two cycles of systems and architecture development took place in Marine for the ANZAC Ship Project. Lessons learned in these Marine implementations influenced development of a solution in Land Division for the M113 Project (described below), and lessons learned from the Land implementation are now being reflected back into the Marine Division for current projects.

A word processing approach based on structured maintenance documents represented the first of the iterated solutions to manage maintenance knowledge. The contract assumed most technical knowledge would be delivered as manually controlled paper documents. Subcontractors would provide most of the technical manuals, however, Marine controlled the support philosophy for the ships, and was thus responsible to produce all maintenance documentation — especially the knowledge intensive "Maintenance Requirement Cards" (MRCs). As detailed by Hall (2001, 2003), MRCs describe maintenance activities to be performed by ship crew or naval on-shore staff. Besides describing how to perform the maintenance, MRCs include technical metadata on what is to be maintained, when the maintenance is to be performed, spare parts, consumables, labor requirements, tools, special and test equipment, anticipated down times, plus a variety of other data needed by maintenance planners and managers. At the outset, it was recognised that key information contained in the documents could be more readily managed if the documents were structured to include fields tagging this information for computer processing.

Beginning in 1992, MRCs were authored and maintained in WordPerfect's merge table format. Different data elements could be entered in semantically specified merge table fields for later selection and formatting by different merge templates. Merge/macro processes automatically extracted particular information required by the logistics analysts (Hall 2001). This saved much labor in its own right, as selected content could be automatically extracted and formatted for particular needs. Formatting changes only required changes to one template rather than editing content in thousands of documents.

Around 1993 the Navy decided to develop a relationally based Asset Management & Planning System (AMPS) to provide computerized scheduling and data collection for maintenance activities. Once details of the AMPS delivery format were agreed, it was comparatively easy to build a macro process and merge templates to extract the required information from the WordPerfect merge tables to produce electronic transfer files. Only templates and associated processes required development. A rudimentary automatic process also validated key data against standard information held in an ILS database. Although there were unsatisfactory aspects to the merge table approach, the value of structured authoring was amply proved (Hall 2001, 2003) for the production of Ship 1 (Australian) MRCs. Maintenance instructions loaded into the AMPS system would then be printed out by the on-board AMPS system to guide maintainers when the maintenance was to be performed.

When authoring of MRCs for Ship 2 (for New Zealand) started it became clear that the "flat" merge table structure used by WordPerfect could not cope with engineering changes that required commonly used key information to be changed across many routines. A single ANZAC ship requires more than 2000 MRCs, and given that no two ships are identical, each new ship required an additional 2000 to be maintained. In some cases a single engineering change might impact content in hundreds, or even thousands of MRCs, each of which had to be manually edited to apply the change. For example, specifying the wrong ID in an MRC for a replacement part could cause a system failure in maintenance. Even with automated validation, it proved impossible to maintain the MRCs completely consistently with changes to the ILS Database, and it was clear that the WordPerfect platform would have to be replaced by something more robust and easier to maintain. Details within MRCs needed to be managed coherently with engineering changes - preferably at a level of structure corresponding to granularity of the engineering data. SGML object management technology offered a solution, as detailed by Hall 2001.

Figure 1 illustrates the potential semantic linkages between elements in structured documents and a product breakdown structure as maintained in a product data, configuration control and engineering change management ("PDM") environment. The heart of an engineering knowledge management system is the PDM engine, which establishes and maintains one or more hierarchical models of the structure of the product. All other data and document content relating to the product can be linked to the specific components ("parts") in product model or breakdown structure (whether this is physical or systems/logical).

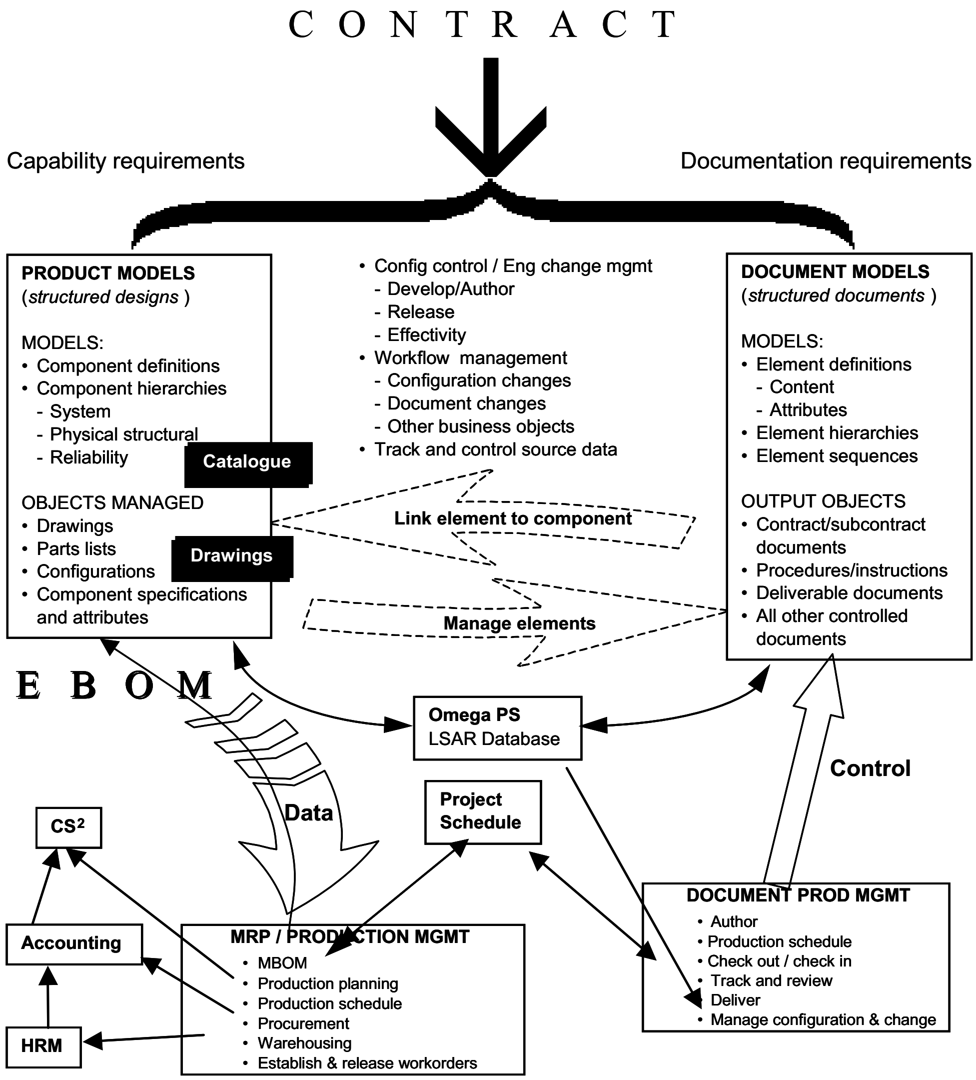


Figure 1. Simplified structure of information, knowledge and control in the fleet production management environment (1997 concept)

Product models are held in a PDM environment. Document models are held in a structured content management environment. Logistic support analysis data is held in an LSAR database. Production control systems include manufacturing resource planning (MRP) and production management, document production management, project scheduling, human resource management (HRM), accounting and Cost and Schedule Control System (CS2). All of this is tracked within a requirements tracking umbrella to ensure contractual requirements are met. The EBOM is the engineering bill of materials for the product. MBOM is the manufacturing bill of materials of the ships as produced.

When new engineering applications were selected in 1998 and 1999, no PDM system recognized document content to a low enough level of structure for document changes to be managed at the same level of granularity as changes to the product model. (PDMs manage document files, but don't see structure within the files). This meant that structured content would have to be managed in a different system from the PDM and the two linked in some way to maintain coherence between engineering and document changes. TeraText, developed by RMIT University (then known as SIM - see http://www.teratext.com.au; Sacks-Davis & Kent 2002), was selected to manage document content. Sherpa Works ("Sherpa" - no longer commercially supported) was selected for the PDM role. Coherence between document content in TeraText and engineering changes in Sherpa was maintained by validating all documents against extracts of key data from Sherpa.

Figure 3 illustrates the system architecture based on Sherpa and TeraText plus other systems forming a complete infrastructure for fleet knowledge management. To collect the operational data to meet the TE&V requirements, Marine developed a software application called OARRS (operational availability recording and reporting system, now known as CSARS) to extract and analyse performance data on the individual ships from the AMPS system. This capability combined with RAN's AMPS and other Tenix systems provide a full lifecycle management capability for the fleet of ANZAC ships.

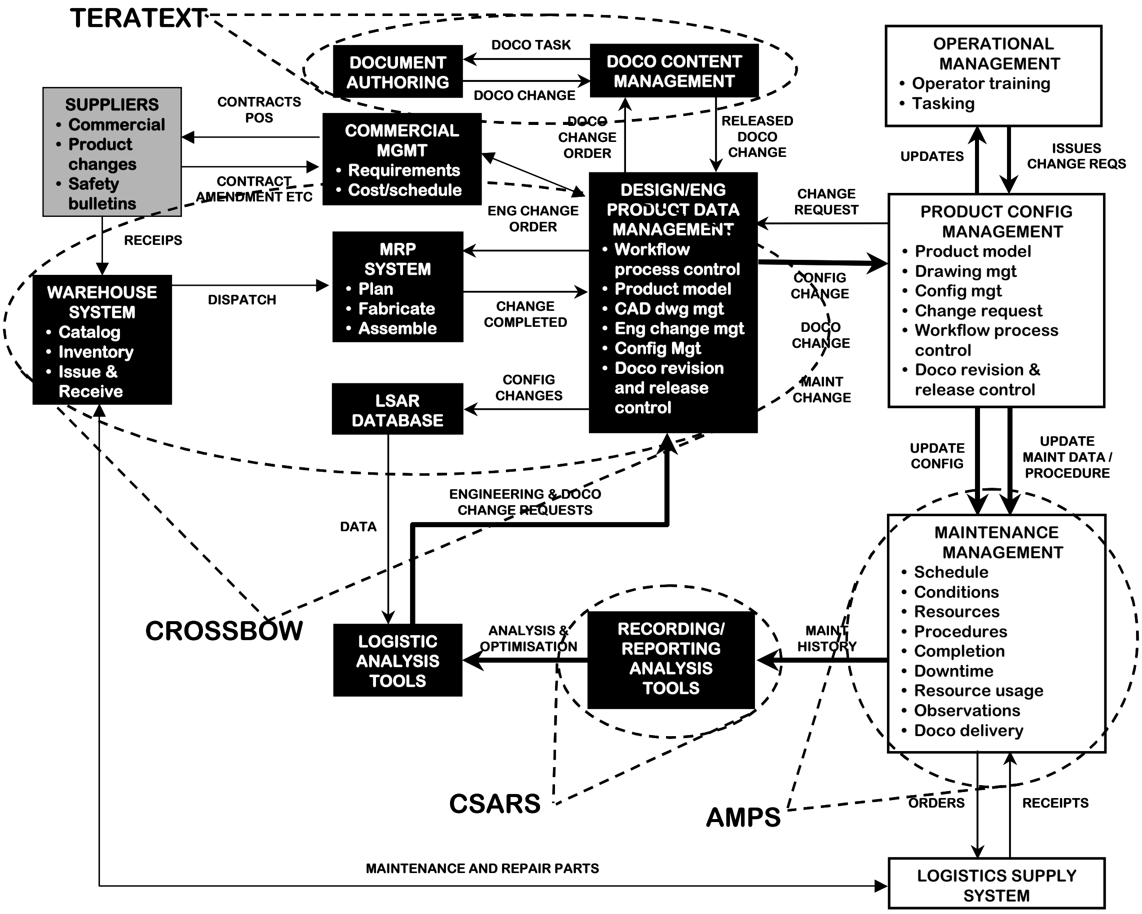


Figure 2. KM architecture for the production and support phases of a fleet project in place by 2002.

The prime contractor's (i.e., Marine) components are shown in black, with white text. The operator's components are shown in white, with black text. Suppliers to the prime contractor are shown in gray. AMPS is the Navy's Asset Management Planning System, CSARS is the Class System Analysis and Reporting System. Crossbow is Marine's data aggregation and normalization engine, and TeraText is the structured content management environment. Sherpa was limited to the Design and Eng Product Data Management role and was also used by the RAN as a configuration management tool. LSAR refers to database applications able to work with logistic support analysis records.

Also, because the ANZAC Project began before integrated enterprise and engineering program management systems existed, Marine developed engineering and ILS technical data in 15 independent software applications (including Sherpa and TeraText). Although text and data were normalized between Sherpa and TeraText, major problems remained to coordinate and normalize key data across many other separate data sources. Tenix addressed this issue by developing a data warehousing and aggregation engine called Crossbow, now marketed commercially (Sykes 2004). Crossbow allowed key technical data to be normalized across all applications and to be visualized in a single user interface (Sykes and Hall 2003; Microsoft Case Studies 2004), for a major cost savings and to eliminate many potential errors in the data that could have surfaced at any time.

Further refinement of architectural concepts for fleet knowledge management then moved to Tenix's Land Division ("Land"). In 2002-3 Land won the project to completely reengineer and upgrade 350 M113 tracked, light armoured vehicles for the Australian Army. The M113 Project was first conceived in 1992 to "upgrade" to the capabilities of a fleet of armored personnel carriers (APCs) first used in Vietnam. As finally agreed, only the bare metal of the original hulls is being reused. Seven major engineering variants will be produced: most hulls will be chopped in two and "stretched", or be even more heavily modified (e.g., to convert what was an armored personnel carrier into a flatbed logistics transport vehicle). Thus, except for parts of the metal hulls, the vehicles are new design and build in every respect (e.g., power trains, weapons systems, electronics, etc.), and require a completely new product technical knowledge package (configuration management, engineering and logistics data, technical, maintenance, and operator's documentation, training materials, etc.). The project is staged, first to build demonstration and "initial production" vehicles for testing and logistics evaluation, with delivery of the first vehicles to an Army company level unit scheduled for 2006. Given that the existing M113's are 30-40 years old, the renewed vehicles are anticipated to have a long life-span in service. Although not required by the project contract, Land's philosophy is that the system for managing technical knowledge should cater for the fact that each vehicle will eventually have its own unique configuration as the result of engineering changes and maintenance activities and that the management system will track the current configuration of each hull.

Land built a highly integrated approach to implement the architecture and data structures illustrated in Figures 2 and 3, as the Configuration Management Information System ("CMIS"). Implementation began late in 2002 with the system in full production by 2004 over all of Land Division's design and production projects. Given that new vehicles are only just beginning to be delivered, the system has not yet been linked directly into the Army's maintenance management environment, although it has been designed with this eventuality in mind. CMIS is based on the Matrix product lifecycle management environment from MatrixOne (see http://www.matrixone.com/matrixonesolutions/index.html; Brouwers 2004). Extending Marine's experience with an earlier version, the latest version of TeraText was used for content management - but in a much more sophisticated integration. In general both Land and the Army operating/maintenance organizations need a PDM capability to know the configuration of the vehicles produced. Land's PDM system maintains three baselines for the product:

- as designed, including the bill of materials (BOM) that guides fabrication of the product within the manufacturing resource planning environment (MRP);
- as built, feeding back details of field changes, parts substitutions, equipment serial numbers, and other information required to be collected in the production process, and
- as delivered — PDM datasets of the physical product(s) as delivered are included in the product technical data deliverables.

The array of product knowledge relating to each component is linked to components in the baselines. The operator's PDM system normally tracks the current configuration state of the product unit as it is maintained in service, and when a component changes, the "where used" approach will quickly retrieve any documentation relating to the change.

CMIS provides a common graphical user interface and electronic workflow environment to control the development, approval, release and publication of all explicit knowledge products for the TLAV project. CMIS is based on MatrixOne's Matrix10 , a complete product lifecycle management system. This provides a graphical workflow environment and master repository able to relate all kinds of business objects to appropriate components in the product breakdown structure; and to electronically progress and control objects through the various business processes of creation, review, revise, signoff, release and fabrication or publication. Through application programming interfaces (APIs), CMIS controls applications in the technical authoring, engineering drawing, logistics analysis and manufacturing areas so all related changes are administered in concert through the one user interface and workflow management environment.

Integration of technical authoring is particularly powerful. To maintain complete integrity of changes to parts and low level elements of document content relating to those parts, document models based on the S1000D standard (TPSMG 2005) are used. TeraText maintains links between particular document elements to specific components in the product breakdown structure in the PDM environment. An engineering change to a part automatically includes all related document elements in the change process. The S1000D standard is specifically designed to facilitate linking document modules to a product breakdown structure — such that a document module relating to a particular component can be reused in a variety of higher level kinds of documents referencing that component. From the standpoint of the end users of the workflow management system, the engineering change process is applied to relevant document components as an integral aspect to changing any other aspect of the information or knowledge linked to a component. Standards for the development of training materials are also being rationalized into the S1000D structure (Shook and Thropp 2004) This ensures that all aspects of knowledge relating to the particular component are considered for change together.

The other aspect of integrating authoring and engineering environments is that it remarkably reduces the risk, effort and cost to perform "impact analyses". A major source of engineering change derives from changes that suppliers make to components they supply or to the data or documentation relating to these components.

In a paper world, when such changes are received, it is a major and risky task to determine what product components or documents may be impacted by the change. Identifying impacts depends on authors' personal knowledge of which parts of which documents are likely to be affected. In general, because supplier changes are often made years after the original documents were written, the authors who wrote the original documents are no longer available. If the analysis misses documents that should be changed, but aren't, this risks an operational failure and whatever consequences cascade from it. Thus, when an impact analysis is performed against a set of paper documents, it may take days or weeks and still be fallible.

By contrast, the TeraText system provides the capability to annotate and link back to the product breakdown structure all component details and source documents used in drafting technical documents (Hall 2003). Impact analyses can be completed automatically in a few minutes to a much higher degree of certainty that all impacts have been identified and assessed. An author still needs to read potentially impacted documents to see whether any changes are required, but this is trivial compared to what is required in the paper world.

Thus, Land Division's CMIS system brings essentially all fleet product related explicit knowledge produced by the supplier organization under a single point of access and control. Controlling all aspects of technical data change, review, and signoff in the common workflow environment, where all associated documentation is linked to the engineering change request/order, ensures all relevant knowledge relating to the change is immediately to hand. CMIS provides the infrastructure to fully support the Technical Regulatory Framework.

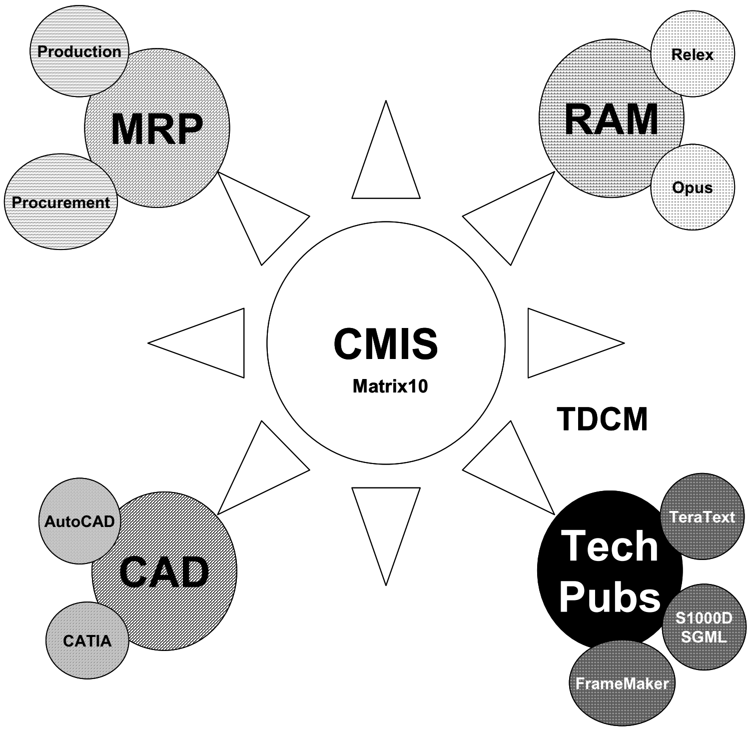


Figure 4.

Fleet knowledge management environment for Land Division. CMIS is the single user interface covering most aspects of the project.RAM includes the logistics tools Opus and Relax for the development of reliability and maintainability information. CAD represents the computer aided design systems, including AutoCAD and CATIA. TechPubs represents the technical authoring environment controlled within TeraText using the S1000D SGML data model and (currently) using FrameMaker as the authoring Tool. MRP represents the manufacturing resource planning environment including procurement and production control.

3 A KNOWLEDGE-BASED ORGANIZATIONAL FORM FOR FLEET MANAGEMENT

Tenix Marine is delivering the last of the ANZAC Ships in 2006, and is now involved in implementing the fourth Tenix iteration of a fleet knowledge management system in a new organizational structure. Responding to the kinds management issues cited by the Auditor General's reports (McNally 1997, McNally and Bates 1999), in mid 2001 the Navy formed a new kind of engineering and support organization for the ANZAC Ships (Bailey 2003). This ANZAC Ship Alliance ("Alliance") is formed from components of the Navy's ANZAC System Program Office ("SPO"), Tenix Marine, and the Combat System Supplier who had been a primary subcontractor to Tenix. The Alliance is a virtual organization formed to manage all changes and upgrades. The aim is to create a high performance collaborative relationship to manage change more effectively than through the use of traditional contracting models.

In the Alliance, Marine provides the primary management systems infrastructure and engineering baseline knowledge. The Combat System Supplier provides its detailed and practical knowledge of the weapons systems and electronics, while the SPO represents the fleet's "owners" and operators. As an organizational entity, the Alliance is responsible to guide and manage engineering and support activities to implement new technologies, optimally maintain capabilities of the ANZAC ships, and to remedy inefficiencies and faults discovered in operation.

An engineering change begins with drafting a change package for circulation to potential 3rd party suppliers for quotations. Installation of a change package is normally subject to open competition. However, better value for money may be gained by allowing the Alliance partners to perform the installation. This decision is based on value for money considerations subject to open book accounting and government probity audits. The Alliance responds to Defence requirements through structured target cost estimating processes. During the estimating process, the Alliance can optimize results by combining features of different competitors' proposals or by trading cost and capability to achieve best mix. The Alliance approach facilitates this optimization while minimizing the ability of any individual commercial partner to unduly influence the outcome (Anonymous 2003). After two years, the Defence Materiel Organization claimed the following benefits to itself and industry as:

- Government's involvement in all Alliance decisions;
- Fewer contract changes, reduced tendering and implementation costs;

- Increased predictability and reliability of forward business; and
- Open communication allowing the strengthening of long term relationships.

The Alliance is currently conducting a "Data Rationalization" project to bring all of the technical knowledge for supporting the fleet under common control within an architecture evolving from that shown in Figure 2. What the alliance is working towards is a more sophisticated and general version of the knowledge testing/improvement cycle illustrated in Hall 2003 (Fig 6) that the Supplier developed during the 10 ship-year test, evaluation and validation phase required under the Warship contract. What is illustrated is an OODA loop (Boyd 1996 - see also http://www.d-n-i.net/fcs/ppt/boyds_ooda_loop.ppt).

- The loop process is seeded with requirements for the operational capabilities to be provided by the fleet product;
- Authors and engineers design and develop a product set that meets the needs and determine knowledge that needs to be provided so the product can be safely and effectively operated and maintained;
- Authors and logistics analysts assemble, write and manage the technical knowledge ("data") package in the document (TeraText DB) and product data (PDM/ILS DB) management systems for internal review and signoff prior to production and delivery to the operators.
- Technical data products are transferred into the AMPS fleet maintenance management system.
- Operators and maintainers use the knowledge embodied in and served up by AMPS in their daily operations and maintenance of the fleet assets.
- Maintainers and operators return observations on results to AMPS, which collects and stores information regarding operations, maintenance, downtimes and any component failures.
- CSARS periodically extracts operational data (observations) from AMPS for logistics and engineering analysis, and provides the analysts with cognitive support highlighting for analysis system components or other factors contributing excessively to operating costs or non-availability of critical systems.
- Helped by CSARS and a variety of logistics analysis tools, logisticians and engineers seek improvements to the engineering of the product or its knowledge base that will resolve the identified problems.
- Recommended changes are circulated for review and signoff by the operator and engineering management, and implemented if approved in the fleet knowledge base.
- Changes to the fleet knowledge base are again downloaded to AMPS and the cycle repeats, to continue through the lifespan of the product,

The Alliance's “Data Rationalization Project”, is currently integrating the architecture illustrated in Figure 1 and Figure 2 within an “umbrella” PDM system like that developed by Land (Figure 3) so all engineering documentation is available to those who need it and so all changes are managed in the one common workflow environment. This is being implemented in the TeamCenter PLM tool offering similar functionality to the Matrix application implemented by Land. Given the proprietary nature of supplier knowledge, Marine and the Alliance will each maintain separate TeamCenter systems, but with the assurance of a seamless interface between the two for those knowledge products needed and used by the Alliance.

We argue that the ANZAC Ship Alliance exemplifies the strategic alliance form of organization (Spekman et al. 1998) based on sharing knowledge and learning (Inkpen 2000) about how to better support the fleet product. Carlile (2004) describes some of the barriers to organizational learning and innovation that boundaries within and between organizations can cause. These take three forms:

- Syntactic difficulties in transferring domain specific knowledge across the boundaries;
- Semantic difficulties in fully understanding and translating the transferred knowledge so it is capable of being used within the new domain; and
- Pragmatic difficulties in evaluating and valuing the effort required to fully transform the knowledge so it can be used in the new domain.

The Alliance's organizational goal is to support the Warship Class using staff and capabilities from three disparate and strongly bounded organizations (two directly competing commercial organizations and a client whose business they compete for) — with the implicit membership of government auditors who represent the ultimate operators and owners of the fleet. Although strategic alliances now represent a well known organizational form, based on a dearth of literature on the topic “support alliance”, this may be a relatively new type of the alliance form. Obvious risks in such an alliance in the fleet management realm would be difficulties establishing process in such a comparatively new organizational form, the potential for impaired engineering and management oversight where processes cross the boundaries of the competing organizations, the propensity to hide errors and mistakes within the partner organization where they occurred, etc. The success of the Alliance depends on the abilities of its partners to break down the boundaries between the disparate organizational groups assigned to carry out alliance activities. It is clear that implementation will benefit from the kind of technological infrastructure described in this paper (Fulk and DeSanctis 1995; Andersen 2004; Huber 1990).

Given that automated workflow is able to progress actions between actors at light speed and effortlessly monitor the progress of actions and activities through the workflows, it should be possible to fully apply the rigorous principles of the Australian Defence Technical Regulatory Framework to engineering change processes within the Support Alliance (and the parent organizations contributing to the alliance). Hopefully, the success of the new organizational form will be demonstrated in Australia, and the spread of its significantly automated cognitive processes to other large engineering projects will be measured by a reduction in the number of engineering disasters around the world.

4 REFERENCES

[1] Adams, D. 2004. Putting and end to the navy blues. Technology/Next, FairfaxDigital, Sydney Morning Herald June 1, 2004 — http://www.smh.com.au/articles/2004/05/31/1085855477337.html?from=storyrhs&oneclick=true

[2] Andersen, T.J. 2004. Integrating decentralized strategy making and strategic planning processes in dynamic environments. Journal of Management Studies 41(8):1271-1299

[3] Anonymous, undated. Quality assurance and the sinking of the largest offshore oil platform — March 2001. — http://www.dlsweb.rmit.edu.au/eng/beng0001/CivilChem/scenic/schooldocs/Oil%20platform%20sinking.ppt.

[4] Anonymous. 2003. ANZAC Alliance — 2 years on. On Target. Defence Materiel Organisation. Oct/Nov 2003. pp. 9-10. — http://www.defence.gov.au/dmo/id/ontarget/edition35/OnTarget35.pdf

[5] Argyris, C. 1999. On Organizational Learning. Blackwell Publishers Limited, Second Ed., Oxford.

[6] Arnold-Moore, T., Fuller, M., Sacks-Davis, R. 2000. System Architectures for Structured Document Data. Markup Languages: Theory and Practice, 2(1):15-44. — http://www.mds.rmit.edu.au/~tja/papers/ml-submitted.pdf

[7] AS 5073(Int)-2003. Knowledge Management. Australian Standard AS 5073(Int)-2003. SAI Global — http://www.standards.com.au/catalogue/script/Details.asp?DocN=AS110152272410

[8] Bea, R. 2003. Quality, reliability and human factors in deepwater drilling & production. Conference Keynote Paper, 21st International Conference on Offshore Mechanics & Arctic Engineering, American Society of Mechanical Engineers, June 2003, Cancun, Mexico, 30 pp. — http://omae.org/acrobat_files/omae03_keynote_speech_bob_bea.pdf

[9] Bailey, F. 2003. The road map for the Defence public private partnership. Keynote Address to Defence Summit 2003, 25 February 2003. — http://www.minister.defence.gov.au/2003/030225.doc

[10] Boyd, J.R. 1976–1996. Unpublished briefings under the name "A Discourse on winning and losing": Introduction (1996), Patterns of conflict (1986), Organic design for command and control (1987), Strategic game of ? and ? (1987), Destruction and creation (1976), and The essence of winning and Losing (1996) available via Defence and the National Interest — http://www.d-n-i.net/second_level/boyd_military.htm

[11] Brown, G. 2002. Decades of disasters: Major project management in defence. Submission to The Senate Committee on Foreign Affairs, Defence and Trade, July 2002. http://www.aph.gov.au/senate/committee/fadt_ctte/completed_inquiries/2002-04/dmo/submissions/sub12.pdf

[12] Brouwers, P. 2004. The CMIS (Configuration Management Information System) solution for the M113 Program at Tenix Defence Systems. CMII World 2004 Conference, Washington DC, August 1 - 4, 2004.

[13] Carlile, P.R. 2004. Transferring, translating, and transforming: An integrative framework for managing knowledge across boundaries. Organization Science 15(5):555-568.

[14] Cohen, B.L. 1990. The Chernobyl accident — Can it happen here?, Chapter 7, The Nuclear Energy Option. Plenum Press — http://www.phyast.pitt.edu/~blc/book/chapter7.html

[15] Commonwealth of Australia 2004. Portfolio budget statements 2004-05, Defence portfolio. — http://www.defence.gov.au/budget/04-05/pbs/2004-2005_Defence_PBS_08_Ch7.pdf

[16] Commonwealth of Australia 2004a. Technical regulation of Army materiel manual (TRAMM). Issue 3.1, Directorate of Technical Regulation — Army — http://www.defence.gov.au/army/PUBS/Tramm/TRAMM.pdf

[17] Commonwealth of Australia 2005. Portfolio additional estimates statements 2004-05, Defence portfolio. — http://www.defence.gov.au/budget/04-05/paes/2004-2005_Defence_PAES_08_ch7.rtf

[18] Dawson, D. and Brooks, B. 1999. The Esso Longford gas plant accident : report of the Longford Royal Commission. Government Printer for the State of Victoria, Melbourne. — [For a summary see http://www.ourcivilisation.com/decline/gasbang.htm]

[19] Dias, R., Valerio, C. 2002. P-36 accident. Presentations from a NPD-seminar 30th April 2002 about the Petrobras 36-ulykken. Petroleumstilsynet, Petroleum Safety Authority Norway — http://www.ptil.no/NR/rdonlyres/76CF1C23-BA45-4CBE-AE9C-AA5F043CFFD1/0/P_36_petrobas1.pdf, http://www.ptil.no/NR/rdonlyres/7A7B068E-750A-40EC-8FF5-07DC4B0449F0/0/P_36_petrobas2.pdf, http://www.ptil.no/NR/rdonlyres/9FA37DE9-BDC6-40F2-AAF9-560F82D7796B/0/P_36_petrobas3.pdf, http://www.ptil.no/NR/rdonlyres/EDE50F0A-FD11-4457-95BC-FD9928307728/0/P_36_petrobas4.pdf, http://www.ptil.no/NR/rdonlyres/88D39CA7-80A3-4D7D-B28E-4179D49A11EA/0/P_36_petrobas5.pdf

[20] Ford, D.N., Sterman, J.D. 2003. The liar's club: concealing rework in concurrent development. Concurrent Engineering: Research and Applications. 11(3):211-219.

[21] Fulk, J., DeSanctis, G. 1995. Electronic communication and changing organizational forms. Organization Science 6(4): 337-349.

[22] Hall, W.P. 2001. Maintenance Procedures for a Class of Warships: Structured Authoring and Content Management. Technical Communication 38(3): 235-247. http://www.tenix.com/PDFLibrary/91.pdf

[23] Hall, W.P. 2003. Managing Maintenance Knowledge in the Context of Large Engineering Projects — Theory and Case Study. Journal of Information and Knowledge Management, 2(2) [Correctly formatted version reprinted in 2(3):1-17]. — http://www.hotkey.net.au/~bill.hall/ManagingMaintKnowledgeinLargeEngiProjects.pdf.

[24] Hope, A.N. 2003. Record of Investigation Into Death. Coroners Act, 1996 Western Australia — http://203.147.135.184/publications/westraliaboi/pdf/HMAS_Westralia_Finding.pdf

[25] Hopkins, A. 2005. The Gretley Coal Mine Disaster: Reflections on the Finding that Mine Managers were to Blame. National Research Centre for OHS Regulation, Working Paper 39 - http://www.ohs.anu.edu.au/publications/pdf/Gretley%20chaps%20AB.pdf

[26] Huber, G.P. 1990. A theory of the effects of advanced information technologies on organizational design, intelligence, and decision making, Academy of Management Review 15(1):, 47-71

[27] Inkpen, A.C. 2000. Learning through joint ventures: A framework of knowledge acquisition. Journal of Management Studies 37(7): 1019-1044

[28] Kemeny, J.G. 1979. Technical staff analysis reports summary: Summary sequence of events. Reports of the technical assessment task force Vol. 1 Staff Reports to the President's Commission on The Accident at Three Mile Island. US Government Printing Office, Washington DC — http://www.threemileisland.org/downloads//195.pdf [Note Dickinson College has placed a comprehensive set of documentation in their Web-based Resource Center relating to Three Mile Island at http://www.threemileisland.org/]

[29] Manderson, J. 2000. Managing technical compliance on large projects. Proceedings, The SETE 2000 Conference Brisbane, Queensland, 15-17 November 2000 — http://www.seecforum.unisa.edu.au/SETE2000/ProceedingsDocs/43P-Manderson.PDF

[30] McIntosh, M.K., Prescott, J.B. 1999. Report to the Minister for Defence on the Collins Class submarine and related matters. Commonwealth of Australia — http://www.defence.gov.au/minister/1999/collins.html

[31] McNally, R. 1997. New Submarine Project. Australian National Audit Office. — http://www.anao.gov.au/WebSite.nsf/Publications/4A256AE90015F69B4A2569060003FAC7

[32] McNally, R., Bates, S. 1999. Management of Major Equipment Projects. Ausralian National Audit Office — http://www.anao.gov.au/WebSite.nsf/Publications/4A256AE90015F69B4A256900000D4D46 .

[33] Mensforth, A. 2004. Technical integrity – achieving it in today's naval environment. Pacific 2004 International Maritime Conference. Sydney, Australia, 3-5 February 2004.

[34] Microsoft Case Studies. 2004. . Australia's Largest Defense Contractor Saves A$11 Million on Major Shipbuilding Project. Microsoft Customer Solution — Case Study — http://www.microsoft.com/resources/casestudies/ShowFile.asp?FileResourceID=5807

[35] Nousala, S., Miles, A., Hall, W.P., Kilpatrick, B. Submitted. Team expertise access maps (TEAM) using mind mapping technologies. Knowledge Management in Asia Pacific, Auckland, 28 — 29 November, 2005. Submitted.

[36] Rios R., 2002, Overview and comments about "Petrobras 36" her history and the accident. Fire & Blast Considerations in the Future Design of Offshore Facilities, June 12-14, 2002 Houston Texas — http://www.fireandblast2002.com/proceedings/presentations/Petrobras%20P36%20Accident%20Investigation%20-%20Ricardo%20de%20Campos%20Rosa%20Rios.pdf

[37] Royal Australian Navy. 2004. Westralia Board of Inquiry. http://www.navy.gov.au/publications/westraliaboi/

[38] Sacks-Davis, R., Kent, A. 2002. TeraText™ database system vs Relational Technology. TeraText Database System White Paper Series. - http://www.teratext.com/documents/TeraTextDBSVersusRelationalTech.pdf

[39] Shook, W.O., Thropp, S.E. 2004. SCORM 2004 Overview. S100D User Group Forum. Clearwater, Florida 5-7 May 2004 — http://www.s1000d.org/content/events/user_forum/clearwater_04/scorm101_s1000dusergroup_v2.pdf

[40] Spekman RE, Forbes TM, Isabella LA, MacAvoy, T.C. 1998. Alliance management: A view from the past and a look to the future. Journal of Management Studies 35 (6): 747-772

[41] Sykes, M. 2004. The need for data integration in a law enforcement environment. Tenix Connections Pty Ltd — http://www.tenix.com/PDFLibrary/225.pdf

[42] Sykes, M; Hall, W.P. 2003. Generating Fleet Support Knowledge from Data and Information. Australian Conference for Knowledge Management & Intelligent Decision Support ACKMIDS 2003 Melbourne, Australia, 11 and 12 December 2003 — http://www.hotkey.net.au/~bill.hall/DataAndInformationInFleetKM(submitted1).pdf

[43] TPSMG. 2005. S1000D: International specification for technical publications utilizing a common source database. Technical Publications Standard Management Group. 2005-05-01 — http://www.s1000d.com/

[44] Woolner, D. 2001. Getting in early: Lessons of the Collins Submarine Program for improved oversight of Defence procurement. Parliament of Australia, Parliamentary Library Research Paper 3 2001-02 — http://www.aph.gov.au/library/pubs/rp/2001-02/02RP03.htm

[45] Yates, A. 2000. Government as an informed buyer: Recognising technical expertise as a crucial factor in the success of engineering contracts. Institution of Engineers, Australia. — http://www.ieaust.org.au/library/publications/govt as informed buyer.doc

APPLICATION OF RANDOM FOREST ALGORITHM IN MACHINE FAULT DIAGNOSIS

Xiao Di Tian Han, Bo-Suk Yang and Soo-Jong Lee

School of Mechanical Eng., Pukyong National University, San 100, Yongdang-dong, Nam-gu, Busan 608-739, Korea.

Abstract: The purpose of this paper is to present a methodology by which rotating machinery faults can be diagnosed. The proposed method is based on random forests algorithm (RF), a novel assemble classifier which builds a large amount of decision trees to improve on the single tree classifier. Although there are several existed techniques for faults diagnosis, such as artificial neural network, support vector machines etc, the research on RF is meaningful and necessary because of its fast executed speed, the characteristic of tree classifier, and high performance in machine faults diagnosis. Evaluation of the RF based method has been demonstrated by a case study on induction motors faults diagnosis. Experimental results indicate the validity and reliability of RF based fault diagnosis methodology. Furthermore, an optimized form of RF is also provided in this paper. We employ the genetic algorithm to strengthen RF. The capability of optimized RF algorithm is proved with same experimental data. It is the evidence that RF based diagnosis methodology can generate more accurate outcome by combining with other optimization method.

Key Words: Random Forests Algorithm, Machine Learning, Genetic Algorithm, Fault Diagnosis, Rotating Machinery

1 INTRODUCTION

The rotating machine plays an important role of modern industry, and is equipped in many crucial departments. This situation causes the problem that its breakdown will result in a huge losing. Therefore, fault diagnosis of machines is gaining importance in industry because of its capability to increase reliability and to decrease possible loss of production due to machine breakdown. Efficient and accurate faults categorized have been critical to machinery operated in normal condition. There are several well-known machine learning methods which also named artificial intelligence, such as artificial neural network (ANN), support vector machines (SVM) etc. The new techniques and their extended research increase the intelligent, preciseness and applicability of diagnosis domain. It exhibits the great potential of combining machine learning methodologies and machinery faults diagnosis theoretic. While the passion of developing machine learning based machinery faults diagnosis methods is increasing, there are a number of obstacles in the presence of researchers. That is, the correct diagnosis of the fault is rather complicated. The reasons are listed as follows:

- Different kinds of faults may result in a certain symptom, or feature extracted from raw data.
- Because of the background noise, some faults are difficult to be distinguished in the machine.
- There are a number of subassemblies with rotating machinery and a high level internal interaction between these subassemblies such as bearings, rotor bar and rotor etc.

Hence, the machine learning based fault diagnosis method which is employed to make hypotheses should be powerful enough to category the malfunctions in a correct way. So that, improving the capability of diagnosis is the motivation to inspirit researchers syncretizing existent technologies and exploring new theories.

In this paper, we introduce and investigate a novel rotating machinery faults diagnosis methodology based on random forests algorithm (RF) [1, 2]. It built a large amount of decision trees out of sub-dataset from a unique original training set by using bagging, acronym of bootstrap aggregating which is a meta-algorithm to improve classification and regression models according to stability and classification accuracy. Bagging also reduces variance and helps to avoid over-fitting. This procedure extracts cases randomly from original training data set and the bootstrap sets are used to construct each of decision trees in the RF. Each tree classifier is named component predictor. The RF makes decision by counting the votes of component predictors on each class and then selecting the winner class in terms of number of votes to it. Details of the theory will be represented in the section two RFs.

Since first introduced by Breiman, RF has been employed in various fields such as astronomy, micro-array analysis and drug discovery and otherwise [3, 4]. RF provides good performance in applications in these fields, and it also can be a competitor for rotating machinery faults diagnosis, because of these distinctive features as below [1]:

- Unexcelled in accuracy among current algorithms
- Runs efficiently on large data bases
- Can estimates of what variables are important in the classification
- Has methods for estimating missing data and maintains accuracy when a large proportion of the data are missing
- Computes proximities between pairs of cases that can be used in clustering, locating outliers, or five interesting views of the data
- Generates an internal unbiased estimate of he generalization error as the forest building progresses

Due to these features and its board application, we investigate the performance of RF based faults diagnosis methodology. Furthermore, an optimized RF methodology is also explored in this paper by combining the genetic algorithm. Experimental result shows optimized RF based methodology achieves a very high accuracy by cooperating RF with GA.

2 RANDOM FOREST AGORITHM

RF which derives from decision tree classifier is an assembled method, it grows tree using CART (acronym of classification and regression trees) methodology to maximum size without pruning. RF has greatly improved classification accuracy resulting from growing an ensemble of trees and making them vote for the most promising class. A convenient method to build the ensembles' random vectors generation via random selection procedure from integrated training set. The constituent in this method is that we prepare k random vectors, Θ_k, which is independent of the past random vectors Θ_1, Θ_2, Θ_3, …, Θ_{k-1} but with the same distribution to build the trees among the RF. Corresponding individual classifier is noted by $C(\mathbf{X}, \Theta_i)$. And then they vote for the most popular class. Breiman names these procedures as random forests. A definition drawn from original paper is available here [1, 2].

Definition 1 A random forest is a classifier consisting of a collection of tree structured classifiers $\{C(\mathbf{X}, \Theta_i), i = 1, \ldots k -1\}$ where the $\{\Theta_i\}$ is independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input vector $\mathbf{X}$.

2.1 Two randomized procedures in RF tree building

As mentioned below, RF enhances classification accuracy significantly compared with decision tree classifier. It is the reason that RF applies two randomized procedures when it builds trees. Each tree is built as follows:

Firstly, assuming that the number of cases in the training set is N and the number of variables in the classifier is M. Select the number of input variables that will be used to determine the decision at a node of the tree. This number, m should be much less than M. Secondly, choose a training set by choosing N samples from the training set with replacement. And then, for each node of the tree randomly select m of the M variables on which to base the decision at that node. Calculate the best split based on these m variables in the training set. Finally, each tree is fully grown and not pruned.

The two distinctive randomized procedures exist among four steps below. That is, RF extracts a fixed quantity from training set randomly, or names it bagging process [5]. RF draws one-third of observations from training set with same underlying distribution. This part of data is named out-of-bag data to get an unbiased estimate of the test set error of an individual tree. The rest of data is used to build the single tree classifier. After the bagging processing, the other randomized procedure is emerged in node splitting while tree classifier is built. Different from normal CART like decision tree splitting algorithm, CART within RF algorithm searches only in m variables which are small amount and drawn at random from all M variables instead of entire variables.

The research of Breiman states why these two randomized procedures make classification accuracy increase effectively. Improvement will occur for unstable procedures where a small change in training set can result in large change between component classifier and classifier trained by entire training set. In RF, whatever the bagging processing or the randomly selection of variables to split the node both make difference in individual tree and forests. Therefore, these two sources of randomness are most important feature of RF.

2.2 Convergence of RF

RF adopts an ensemble of decision trees and determines the categorical classes by majority vote algorithm. Thus a serious consideration of over-fitting is necessary to test the performance of RF. Normally an over-fitting will occur where learning is performed too long or where training examples are rare, the learner may be limited in very specific random features of the training data that has no causal relation with target function. But RF can avoid the over-fitting completely, which was proved in Refs [1]. To affirm this point, we define a margin function at first.

Given an ensemble of a series of classifiers $C_1(\mathbf{X})$, $C_2(\mathbf{X})$, …, $C_k(\mathbf{X})$, and with the training set drawn at random from the distribution of the random vector Y, $\mathbf{X}$, define the margin function as

$$mg(\mathbf{X},Y) = av_k \, \mathrm{I}(C_k(\mathbf{X}) = Y) - \max_{j \neq Y} av_k \, \mathrm{I}(C_k(\mathbf{X}) = j) \quad (2)$$

Where $\mathbf{X}$ is the input metric, av_k is the average number of votes at $\mathbf{X}$, Y for the right class and I( ) is the indicator function. The margin measures the extent to which the av_k exceeds the average vote for any other class. The larger in the margin, the more confidence in the classification.

According to this function, the generalization error is given by:

$$PE^* = P_{X,Y}\left(mg(\mathbf{X},Y) < 0\right) \quad (3)$$

Where $P_{\mathrm{X},Y}$ indicates the probability which is over the $\mathbf{X}$, Y space.

Theorem 1 As the number of trees increases, for almost surely all sequences Θ_1, PE^* converges to:

$$P_{\mathbf{X},Y}\left(P_\Theta\left(C(\mathbf{X},\Theta) = Y\right)\right) - \max_{j \neq Y} P_\Theta\left(\left(C(\mathbf{X},\Theta) = j\right) < 0\right) \quad (4)$$

Theorem 1 is proved with the strong law of large numbers and the tree structure. It indicates that it is unnecessary for RF to employ common anti-overfitting methods for instance, cross-validation, early stopping, etc. RF do not overfit when more trees are added, meanwhile it result in a limiting value of the generalization error. This is another important feature of RF beside the two randomized procedure mentioned above.

2.3 Accuracy of RF depending on strength and correlation

In last section, the anti-overfitting characteristic of RF is proved. But we concern more about its accuracy. According to the analysis built in references, an upper bound of RF can be derived for the generalization error in terms of two parameters that are measures of how accurate the individual classifiers are and the dependence between them. These also lead an in-depth view of how RF works. Firstly we define a margin function and raw margin function for RF.

The margin function for a random forest is:

$$mr(\mathrm{X},Y) = P_\Theta\left(C(\mathbf{X},\Theta) = Y\right) - \max_{j \neq Y} P_\Theta\left(C(\mathbf{X},\Theta) = j\right) \quad (5)$$

The raw margin function is:

$$rmg(\Theta,\mathbf{X},Y) = I\left(C(\mathbf{X},\Theta) = Y\right) - I\left(C(\mathbf{X},\Theta) = \hat{j}(\mathbf{X},Y)\right) \quad (6)$$

Distinctively, $mr(\mathbf{X}, Y)$ is the expectation of $rmg(\Theta, \mathbf{X}, Y)$ with respect to Θ. And the strength of the number of individual classifiers $\{C(\mathbf{X}, \Theta)\}$ is:

$$S = E_{X,Y} mr(\mathbf{X},Y)$$

where $E_{\mathbf{X},Y}$ is the expected value of margin function over $\mathbf{X}$, Y space. Then we compute the variance of margin function:

$$\mathrm{var}(mr) = \bar{\rho}\left(E_\Theta sd(\Theta)\right)^2 \le \bar{\rho} E_\Theta \,\mathrm{var}(\Theta) \quad (7)$$

Write

$$\mathrm{E}_\Theta \,\mathrm{var}(\Theta) \le \mathrm{E}_\Theta\left(\mathrm{E}_{\mathbf{X},Y}\, rmg(\Theta,\mathbf{X},Y)\right)^2 - S^2 \le 1 - S^2 \quad (8)$$

Where $\bar{\rho}$ is the mean value of the correlation, $sd(\cdot)$ is standard deviation of $rmg(\Theta,\mathbf{X},Y)$. Considering equations (7), (8) and Chebychev inequality, theorem 2 can be concluded.

Theorem 2 An upper bound for the generalization error is given by

$$PE^* \le \bar{\rho}\left(1-S^2\right)/S^2 \tag{9}$$

Although the bound is likely to be loose, it full fills the same suggestive function for random forest as VC-type bounds do for other types of classifiers. It shows that the two ingredients involved in the generalization error for random forests are the strength of the individual classifiers in the forest, and the correlation between them in terms of the raw margin functions. There is a conclusion drawn from this upper bound, that is the smaller this ration is, the better performance RF provided.

3 GENETIC ALGORITHM

RF is strengthened by a standard genetic algorithm (GAs) [9] in this paper. GA is a simulation of evolution where the rule of survival of the fittest is applied to a population of individuals, or it can be considered as a parallel searches procedure that simulates the evolutionary process by applying genetic operators. Comparing with other search algorithms, GA has been well-known for its superior performance. And a most powerful feature of GAs is the great simplicity of it. They do not need too much code and no differentiability or continuity requirements to be satisfied. The usual GA flowchart and steps are shown as follow:

Step 1: Coding, generate an initial population (usually a randomly string)

Step 2: Fitness evaluation, apply some function or formula to the individuals to get the fitness of each individual.

Step 3: Selection, according to the fitness, individuals are selected to be the parents of next generation.

Step 4: Crossover, it is used to create two child individuals from the parent which pass the selection successfully via exchanging their chromosomes.

Step 5: Mutation, it assigns a new value to a randomly chosen gene and is controlled by a mutation probability.

Step 6: Repeat steps 3 to 5 until the evolved result satisfy the termination criteria, or a certain fixed number of generations are achieved.

4 EXPERIMENT PLATFORM AND MOTOR FAULTS DATA DESCRIPTION

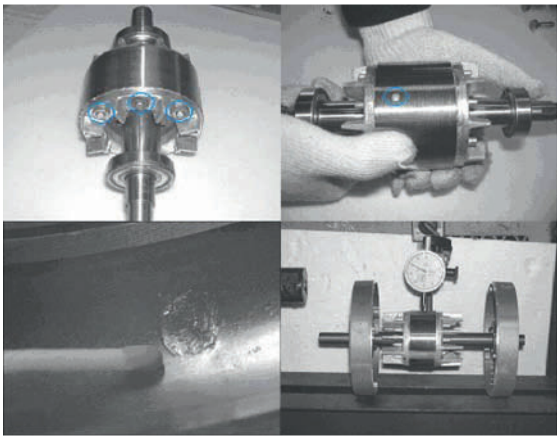


Figure 1. Faults on introduction motor.

The experiments are designed to simulate six most universal categories of induction motors faults which are broken rotor bar, bowed rotor, bearing outer race fault, rotor unbalance, adjustable eccentricity motor (misalignment) and phase unbalance, first four motor faults are shown in Fig. 1 as examples. The load of the motors can be changed by adjusting the blade pitch angle or the number of the blades. The platform of these experiments consists of six 0.5kW, 60Hz, 4-pole induction motors, pulleys, belt, shaft and fan with changeable blade pitch angle.

Three AC current probes and another three accelerometers were used to measure the stator current of three phase power supply and vibration signals of horizontal, vertical and axial directions for evaluating the RF based fault diagnosis system.

After measuring the raw data, a preprocessing and feature extraction are implemented on the data to obtain the most important features for the RF based diagnosis methodology [6]. Finally, there are 63 features left which are prepared for the next procedure, induction motor faults diagnosis by RF.

5 INDUCTION MOTOR FAULTS DIAGNOSIS

In this section, RF was run on the induction motor faults data. The experimental results are given in Table 1. Confusion matrixes for the training data in RF are given by Table 2 which shows the accuracies of each fault class for testing data with 1200 trees and selecting 1 variable every split.

Table 1

Faults diagnosis accuracies based on RF

Trees No.	Split variables	Test accuracy (%)	Trees No.	Split variables	Test accuracy (%)
200	1	88.89	2000	5	73.34
500	1	94.44	5000	5	72.23
1200	1	95.56	10000	5	74.44
2000	1	93.33	200	8	81.22
5000	1	92.23	500	8	83.33
10000	1	92.25	1200	8	82.34
200	5	71.11	2000	8	83.33
500	5	75.56	5000	8	78.89
1200	5	72.23	10000	8	77.78

The experimental results in Table 1 show three characteristics of RF clearly:

First, compared with the number of component classification trees, the parameter, random split number at each node, is more sensitive to the classification accuracy. Hence a prudential searching procedure is necessary to find the best split variables number by an experimental way. Second, if the split variables number is decided, the sum of individual tree classifiers should achieve appropriate quantity to get a better performance. Last one, when we increase trees into a high number, for example 5000 or 10000, there is no over-fitting occurred but a little undulate existed.

Table 2 indicates that incorrect diagnosis of RF based methodology often occurs at certain fault category. So we can apply some assistant diagnosis methods which are function in that specific kind of fault to improve the diagnosis precision. Additional, variable importance is another strongpoint of RF. But in this case, using 18 most important features replacing the all just can get an accurate rate 93.32%, slightly lower than the best correct classification rate. The reason is that replacing all by the important features is able to increase the accuracy only when the data is adulterated with noise at a certain extent.

Table 2

Accuracy of each fault class for test data with 1200 trees and selecting 1 variable every split

Class No.	1	2	3	4	5	6	7	8	9	Accuracy (%)
1	10	0	0	0	0	0	0	0	0	100
2	0	10	0	0	0	0	0	0	0	100
3	0	0	10	0	0	0	0	0	0	100
4	0	0	0	10	0	0	0	0	0	100
5	0	0	0	0	7	3	0	0	0	70
6	0	0	0	0	0	10	0	0	0	100
7	0	0	0	0	0	0	10	0	0	100
8	0	0	0	0	0	1	0	9	0	90
9	0	0	0	0	0	0	0	0	10	100

In general, the normal RF has achieved the satisfied fault diagnosis accuracy. But it should be noticed that two parameters, the number of trees and random split number, which greatly affect classification result are set manually. It means accuracy of normal RF depends on researcher's experience. This situation exists at almost all the applications of RF. So that applying the genetic algorithm to do the parameter optimization is exigent. The effect of this cooperation is proved by using same data. According to the pervious research, the number of trees and random split number are limited in the range from 500 to 1500 and from 1to 10 respectively. In order to reduce executed time of GA program and find the optimized point synchronously.

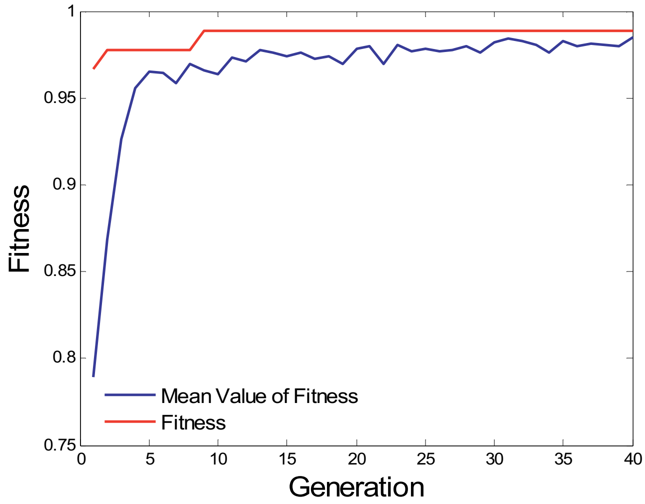


Figure 2. Optimization trace within 40th generation.

Figure 2 shows the trace information of every generation. Fitness adopts the classification accuracy of test data set. The upper curve is the best fitness value and the other one is mean fitness value of each generation. The risen trend of mean fitness value convergent indicates that GA well cooperates with RF based methodology on the motor fault diagnosis, and best fitness value lays out the optimization point which is 908 trees and 1 random split created by 9th generation. The classification accuracy at this point achieves the 98.89%, 3.33% higher than the best value of normal RF. It means GA can enhance the capability of RF algorithm distinctly.

6 CONCLUSIONS

In this paper, the RF and optimized RF based faults diagnosis methodology of rotating machinery were investigated. The performance of two methodologies was proved by the faults diagnosis test of an induction motor. The optimized approach attains a high correct rate of diagnosis, 98.89%. It means optimized RF based method is competitive with other classification method. Especially, the assemble classification trees method and even faster then some of them [1, 7], it is proved by other multi-classes classification applications.

The extended research will focus on decreasing the redundancy of the RF and try other optimization algorithm or more effective voting principle.

7 REFERENCES

[1] Breiman L. (2001) Random forests. Machine Learning, 40(1), 5-32

[2] Breiman L., Random Forest User Notes.: ftp://ftp.stat.berkeley.edu/pub/users/breiman/notes_on_random_forests_v2.pdf

[3] Gislason PO, Benediktsson JA & Sveinsson JR (2004) Random forest classification of multisource remote sensing and geographic data. Proceedings of IGARSS ’04 IEEE International, 2, 1049-1052

[4] Remlinger K (2003) Introduction and application of random forest on high throughput screening data from drug discovery. Proceedings of Workshop for the SAMSI Program on Data Mining and Machine Learning

[5] L. Breiman L (1996) Bagging predictors. Machine Learning, 24, 123-140

[6] Yang BS, Han T & An JL (2004) ART-KOHONEN neural network for fault diagnosis of rotating machinery. Mechanical Systems and Signal Processing, 18, 645-657

[7] Pal M (2003) Random forests for land cover classification. Geoscience and Remote Sensing Symposium, IGARSS

[8] Dargupta H & Dutta H (2004) Orthogonal decision trees. Proceedings of 4th IEEE International Conference on Data Mining, 427-430

[9] Goldberg GE (1989) Genetic Algorithm in Search, Optimization and Machine Learning. Addison Wesley, New York

Acknowledgements

This work was partially supported by the NURI Project in 2005.

EXPLOITING TECHNOLOGY FOR REAL TIME HIGH VOLTAGE NETWORK FAULT MANAGEMENT

Randall Jones and Brett Handley

Powerlink Queensland, PO Box 1193, Virginia, Queensland, 4014, Australia

Abstract: Changes in the electricity supply industry have led to transmission companies being exposed to increased pressure associated with reliability and quality of supply issues. The effectiveness of the fault management process is fundamental in mitigating business risks associated with the impact of faults on the high voltage network. The existence of large volumes of data being captured and stored by intelligent field devices associated with almost all items of high voltage plant and equipment, has resulted in the root cause of network events being able to be readily identified in a post fault timeframe. However, if this data can be accessed remotely, and most importantly turned into "information", in the "real time" fault timeframe, it can be used to make informed decisions about the suitability of faulted plant for service. Using this philosophy, Powerlink Queensland has developed and implemented an innovative fault management strategy, aimed at realising the business goals. This paper outlines Powerlink Queensland's business approach to fault management. In particular, the paper focuses on the key elements of the strategy that have enabled the business goals to be realised. A case study provides an example of the benefits being realised by this innovative approach.

Keywords: Fault Management; Asset Monitoring; HV Networks

1 INTRODUCTION

Powerlink Queensland is a government owned corporation that owns, operates and maintains Queenslands high voltage electricity transmission network. The network extends 1700km from north of Cairns to the New South Wales border – approximately half of Australia's eastern seaboard.

It is common knowledge the electricity supply industry has undergone significant restructuring in the face of deregulation in recent years [1]. Modern power systems have become more complex to operate and control, and transmission infrastructure is being run to the limits of its technical envelope to achieve the demanding performance requirements of Transmission Network Service Providers (TNSPs).

These ongoing changes have forced TNSPs to think outside the square when it comes to utilisation of asset capability.

Powerlink as a regulated electricity transmission company faces a continual challenge to reduce system operation and maintenance costs while increasing network availability and reliability. [5]

The introduction of the Australian Competition and Consumer Commission (ACCC) transmission service standards, and the associated financial penalties and bonuses for transmission companies, have increased the need to be able to make informed decisions about the health of high voltage network plant and equipment.

The key, to allow informed decisions to be made about the suitability of service of high voltage network plant and equipment following an event, is the availability of accurate information.

This paper outlines the ways in which Powerlink Queensland is exploiting the capabilities of equipment technology to obtain this accurate information, to strive for operational excellence in transmission service provision.

2 THE NEED FOR INFORMATION

In November 2003 the Australian Competition and Consumer Commission (ACCC) released transmission service standards guidelines [2]. These standards have had a marked impact on the way TNSPs view network operational performance. According to Powerlink Queensland Chief Executive, Gordon Jardine, being operationally 'just OK' is no longer acceptable for grid customers and other stakeholders. 'Operational Excellence' has since become a key Powerlink business objective with financial penalties imposed upon TNSPs failing to meet the ACCC service standards. [4] There are also monetary incentives offered to TNSPs exceeding the standards.

Financial incentives under the transmission service standards guidelines are attributed according to delivery against predefined operational measures. These include:

- Measure 1 – Transmission Circuit availability;
- Measure 2 – Average outage duration; and
- Measure 3 – Loss of supply event frequency index.

Unfortunately, network incidents will often occur due to influences outside the TNSPs' control. The circuit availability measure, loss of supply frequency index, average outage duration and network constraints are impacted according to the success of the corporate fault management strategy. The success of the strategy hinges on critical accurate information being available, to allow restoration plans to be formulated and resources to be deployed, to reinstate the affected network in the most effective and efficient manner.

3 AVAILABILITY OF DATA

3.1 Plant and Equipment

In the late 1990s, a period of intense capital investment and expansion commenced on Powerlink's transmission network. It was quickly identified that high voltage plant and equipment manufacturers were embracing technology, and supplying products that came standard with intelligent monitoring systems capable of providing data about the health of the device, both in a real-time timeframe and as stored historical trends and records. These monitoring systems were being applied equally to substation plant, such as circuit breakers and transformers, as well as secondary systems including protection, control, and communications systems. As an example, Powerlink has high voltage circuit breakers in service that each host a total of twenty-one intelligent devices on board (seven per phase) performing state of the art monitoring and data provision functions.

This monitoring data generally lives on the item of plant, in an intelligent device, or in the case of secondary systems, within the device itself.

This monitoring data is very detailed and as such not suitable for transmission to the Substation Control System (SCS) or transmission via Supervisory Control and Data Acquisition (SCADA) to Powerlink's statewide network control centre (NCC).

Generally, manufacturers intended for the information to be obtained and analysed by service personnel on site, trained in the use of extraction tools such as laptop computers and trained in the expert assessment of the data, in order to turn it into information.

3.2 Substation Control Systems

In the early to mid 1990s, Powerlink began implementing computer based distributed control system technology as its substation control system (SCS). The SCS was implemented with a Local Control Facility (LCF) that was both a Human-Machine Interface (HMI) and an event capturing and logging system. The event log of the site LCF provides the lowest level detail of substation event information and is crucial to the post fault analysis function of system events.

In the early 2000s, a new and improved substation integrated secondary system was designed by Powerlink to take full advantage of the remote access technologies and infrastructure being implemented. The secondary system provides advanced automation and condition-monitoring facilities, including significant fault diagnosis and alarming features. [6]

4 OBTAINING DATA FROM THE FIELD

With consideration of the business drivers, it very quickly became apparent, and increasingly justifiable, that there were some big advantages to be realised by remotely accessing data that existed at the remote site.

As part of the capital expansion in the late 1990s and continuing into the 2000s, Powerlink's telecommunications infrastructure has been significantly upgraded and expanded. This was necessary to meet the requirements of the National Electricity Rules in the areas of protection and SCADA.

High speed digital communication existed between Powerlinks Brisbane office and remote substations. The site data was available and waiting to be obtained, and a communications infrastructure back to the office was in place. The next step to provide a connection from the intelligent devices to the communications infrastructure was easily justifiable when considerations of the advantages to be obtained were completed.

Powerlink established an asset monitoring strategy to reduce operational and maintenance costs, while seeking to reduce the frequency and duration of system outages. A key aspect of this strategy was the remote monitoring of high voltage equipment to enable rapid diagnosis of incipient failure trends and faults prior to a system or equipment outage. [5]

From this strategy, the justification, design and implementation of Powerlink's Operational Wide Area Network (OpsWAN) for remote monitoring and interrogation of plant and equipment was borne.

The OpsWAN system comprises a number of strategically located computer servers that provide the gateway to plant and equipment located in that geographic area. These servers communicate via an operational Wide Area Network (WAN) comprising of 2 Megabit communication channels.

Powerlink currently has 70 substations remotely accessible from the OpsWAN system, providing remote access to over 1500 intelligent primary plant and secondary system devices from its Brisbane office. Devices connected include:

- Video camera systems;
- Protection relays;
- Battery systems;
- Telecommunications equipment;
- Substation local control facilities;
- Discreet device local control facilities
- Hydran transformer gas monitoring units;
- Power transformer management systems;
- Circuit breaker monitoring systems; and
- Other plant specific plant monitoring systems.

A typical configuration of an OpsWAN installation is shown in Figure 1.

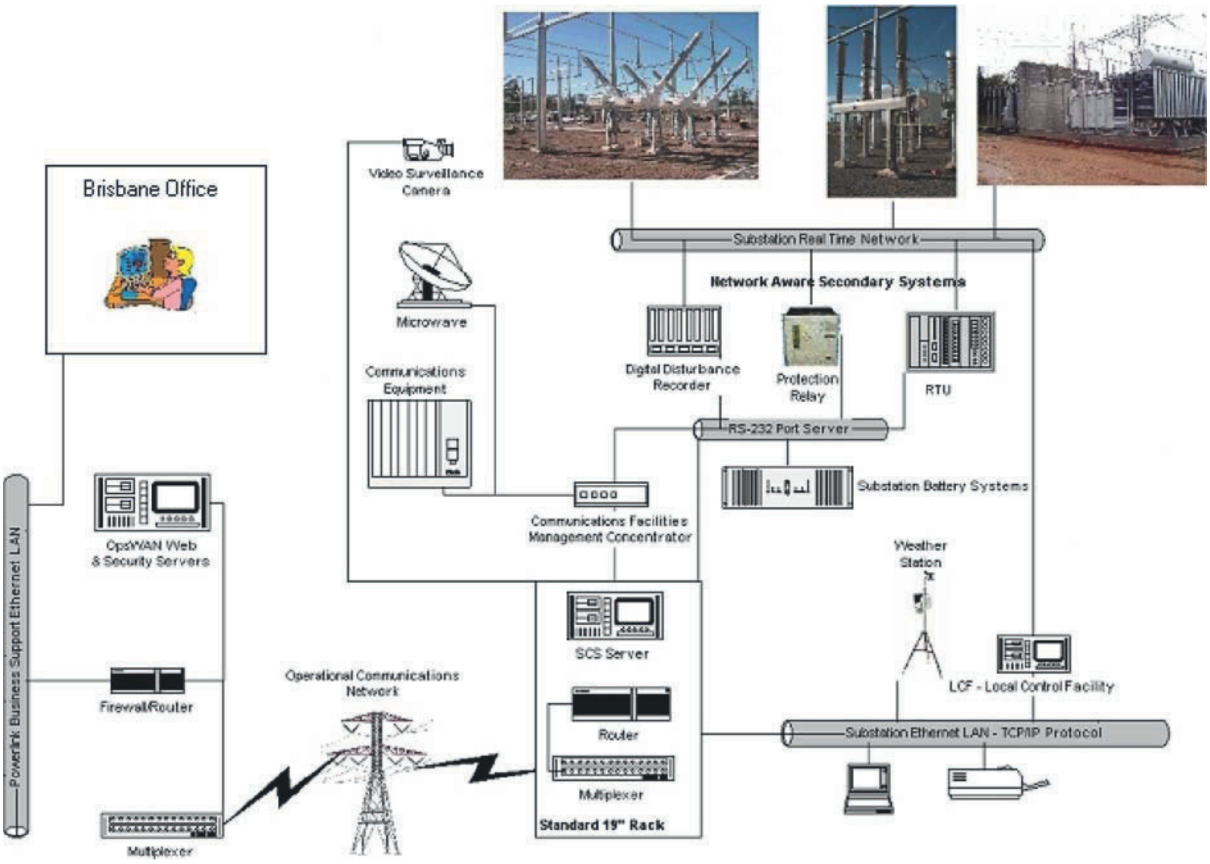


Figure 1. Typical configuration of an OpsWAN installation

5 TURNING DATA INTO INFORMATION

The next part of Powerlink's asset monitoring strategy is to turn this remotely obtained low-level detailed data into value added accurate information. It is this value added accurate information that is critical, to allow informed decisions to be made about the suitability of service of high voltage network plant and equipment following an event.

As part of Powerlink's Asset Monitoring Strategy, an Asset Monitoring Team (AMT) was formed. The AMT has three key objectives associated with the operation and maintenance of Powerlink's high voltage transmission network. These are to reduce costs, minimise exposure to potential liability, and ensure quality of supply. To achieve these objectives, one of the team's tasks is to turn the copious amounts of field data, now available in the office, into value added accurate information in a timeframe suitable for use in real time system event restoration and action plans.

An innovative part of Powerlink's Asset Monitoring Strategy entailed the design of a high level framework for the provision of information from the analysis of raw data.

An overview of the Powerlink Data to Information Framework is shown in Figure 2.

Within the framework, the upper layers draw on and are dependent on the lower ones. For example, prior to data interpretation being done, some level of data management is required. In turn, data management requires an emphasis on communications systems and data capture capability. [3]

The seven layer data-to-information pyramid below summarises this process.

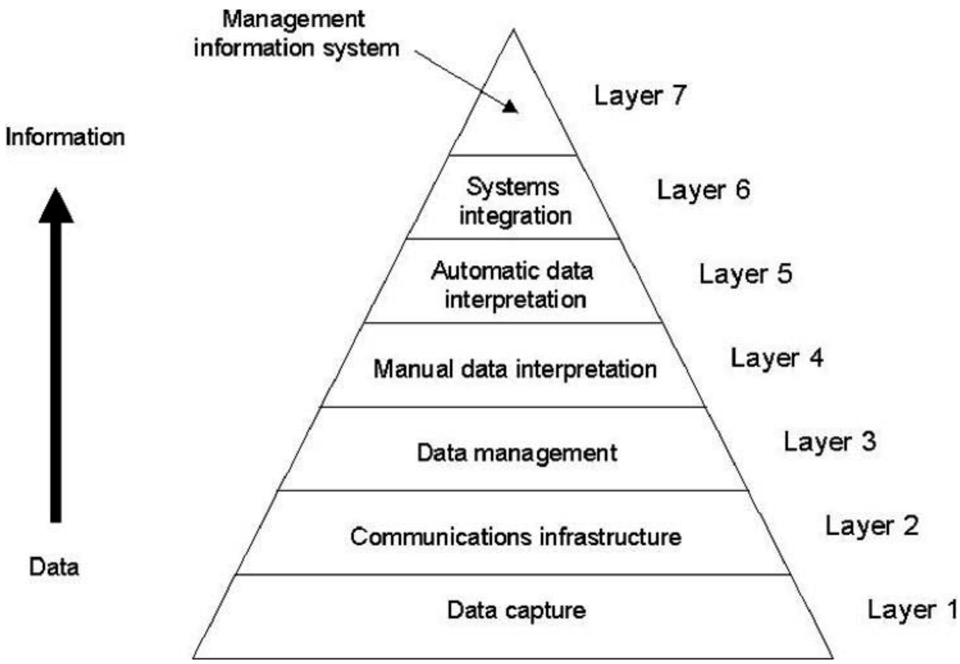


Figure 2 – Powerlink Queensland Data-to-Information Framework

The ability to obtain site data remotely from the office and turn it into value added information in a real time timeframe was achieved. The next step is to integrate this capability into the Fault Management process.

Powerlink's Asset Monitoring Strategy had also identified that changes to the fault management process were required. A review of the fault management process was undertaken and changes made as discussed later.

6 FAULT MANAGEMENT PROCESS REVIEW

To Powerlink's management, it was clear the ability to obtain detailed site data and turn it into valuable information quickly would provide significant benefits from a fault response perspective. The creation of a new role to cover this requirement would also provide significant benefits in exploiting the available data from a proactive condition and performance monitoring perspective. A business case put forward to the Powerlink Executive resulted in the approval and successive formation of the Asset Monitoring Team (AMT) to perform a number of new functions, including fault management. [4]

The AMT needed to interface efficiently with the traditional functions of fault management. The conventional fault management process was modified such that the Network Control Centre (NCC) continued detection of plant and equipment faults via the SCADA system, but then handed over the fault management responsibility to the AMT.

The NCC monitors the network, and, if a fault occurs, transfer the fault management responsibility to the AMT. The AMT are available on-call 24 hours a day, 7 days a week. This single point of contact ensures there is no confusion as to who the NCC controller should call. This hand over process allows the NCC operator to concentrate on their core task of monitoring and operating the high voltage network.

A flowchart of the new fault management process is shown in Figure 3.

The AMT then obtains the detailed data from the appropriate substations via OpsWAN and analyses the data. Once the cause of the fault has been established, the AMT can then assign a priority to the fault and request the Maintenance Service Provider (MSP) attend to the fault immediately, next day, next working day or next opportunity, depending on the urgency of the fault and based on policies set by Powerlink's Asset Manager. A further benefit of this strategy is that the cause of the fault is known up front, enabling appropriately skilled staff with the necessary tools and spares to be sent to site, hence avoiding multiple site visits. In certain circumstances, where technology allows, it is possible for the fault to be repaired remotely avoiding the requirement for MSP involvement.

There are feedback loops between the different parties to ensure everyone is kept abreast of developments as necessary. The AMT also fulfils the coordination role between the various stakeholders.

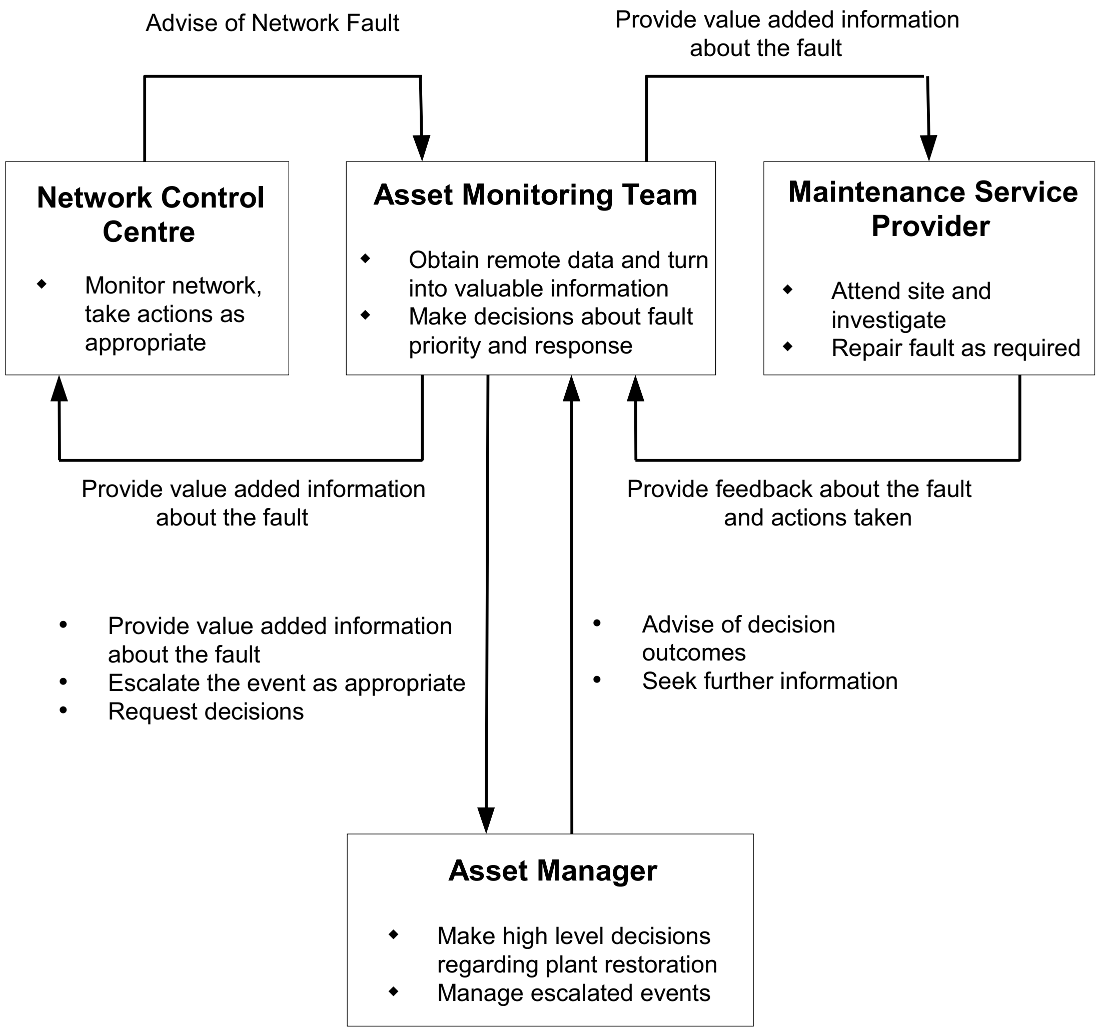


Figure 3. Powerlink's New Fault Management Process

As a result of the new fault management process, the AMT is involved with each fault that occurs on Powerlink's high voltage network. As the AMT is intimately involved in the fault management decision making process their knowledge and experience is made available for every incident. This involvement is also beneficial in post fault analysis to ensure the same fault does not occur again, and in looking for failure trends across the network.

On reflection, the traditional fault management process does not cater for the existence of the AMT. The AMT is necessary to exploit fully the technology and communication capability available. The NCC cannot afford to be distracted from their core business of operating the network, therefore it was decided not to extend the roles of the NCC controllers.

7 ADVANTAGES OF THE NEW ASSET MONITORING STRATEGY

From the perspective of the Asset Manager, MSPs and NCC controllers, significant improvements are being realised in the amount of valuable information being made available. In particular, benefits being achieved by the new process are the:

- application of the Asset Manager policy and procedures in a consistent manner;
- application of the fault knowledge obtained from one component failure across the entire network;
- exploitation of the technology and related infrastructure (this is clearly demonstrated in the following case studies);
- prioritisation of the fault response, so MSPs are not called out unnecessarily;
- realisation that the AMT has the knowledge to perform proactive as well as reactive fault management.

Deviating from traditional fault management practices places Powerlink in a better position to utilise installed technologies to meet the changing business and regulatory body objectives. The combination of technology and efficient process has had a marked impact on the speed of response to power system faults and is progressively eliminating recurring faults through better monitoring and works management. [4]

The combination of these benefits results in reduced cost of fault management activities. Based on operational experience to date using the new fault management processes and the technology available, Powerlink has reduced the number of fault related site visits, to sites connected to OpsWAN, by typically 20% and reduced the requirement for immediate call outs by 50%.

8 CASE STUDIES

8.1 Case Study 1 – Loss of Sulphur Hexaflouride (SF6) insulating gas from a Circuit Breaker

The following case study provides an example of two real network events. Both events occurred at a critical 275/132/66kV transmission substation providing point of supply for distribution company load as well as critical mining infrastructure load. This substation has a modern Substation Control System (SCS) with a Local Control Facility (LCF) that provides low-level detail about the root cause of any alarms initiated from the site. Summary alarms are sent to the Network Control Centre (NCC) to advise of loss of functionality of equipment on site, or priority of attendance required to an equipment fault in line with Powerlink's Asset Manager policies. This substation is located approximately 4 hours drive from the primary maintenance service provider's depot.

Event 1 occurred in October 2003 and was managed under the previous (traditional) fault management model. This event has been provided to allow comparison to Event 2, which was associated with the same circuit breaker at the same substation in October 2005.

8.1.1 Event 1 – October 2003

This event occurred under the traditional fault management model.

At 1440hrs several alarms were received at the Powerlink NCC. The alarms advised of both an urgent and non-urgent condition associated with a 132kV Hybrid circuit breaker. There was no other information available about the alarms from site. The maintenance service provider responsible for this site was advised, with the estimated time of arrival being approximately 1830hrs.

At 1717hrs a protection trip of one of the two 132kV busbars occurred. The busbar that tripped was the one connected to the hybrid circuit breaker that was alarming. Maintenance service provider staff attended site and it was subsequently determined that the cause of the urgent alarm from the 132kV Hybrid circuit breaker was due to a rapid loss of SF6 insulating gas. This was caused by a major leak associated with one of the rupture disks. An internal flashover subsequently occurred when the SF6 insulating gas density could no longer provide adequate insulation for the system voltage. Significant damage occurred to the pole resulting in the need for it to be sent back to the manufacturer for repair.

8.1.2 Event 2 – October 2005

This event occurred under Powerlink's new fault management policy, which incorporates the use of technology to remotely access substation low-level detail data, and a process incorporating the AMT to turn this data into information is a real time timeframe.

At 1543hrs several alarms were received at the Powerlink NCC. The alarms advised of both an urgent and non-urgent condition associated with a 132kV hybrid circuit breaker.

At 1545hrs, the NCC operator called the AMT to advise of the alarms and hand over the fault management responsibility.

AMT staff used OpsWAN to remotely access the site LCF, the device that contains the low-level detailed data associated with the alarms received by the NCC.

At 1554hrs, after remotely accessing the substation LCF, AMT staff identified the root cause of the alarms as a rapid loss of SF6 insulating gas from a 132kV hybrid circuit breaker.

At 1555hrs, AMT staff advised the NCC operator that urgent de-energising of the affected circuit breaker was required to avoid internal circuit breaker flashover, in turn causing a protection trip of one of the two 132kV busbars.

The NCC staff immediately developed a plan to shift load and, with the assistance of monitoring information from the AMT on estimated time to flashover, had remotely de-energised the affected 132kV hybrid circuit breaker by 1639hrs.

No load was lost for the event, a costly hybrid circuit breaker internal flashover was avoided and network constraints and loading issues were significantly reduced compared to Event 1 in October 2003.

8.2 Case Study 2 – Trip of two 132kV Busbars at a critical Central Queensland substation.

This case study steps through the events associated with a forced outage at one of Powerlink's 132kV substations.

The event commenced with the Bus Zone protection trip of two out of the four 132kV busbars. The single line diagram is shown in figure 4. This event occurred outside of normal business hours.

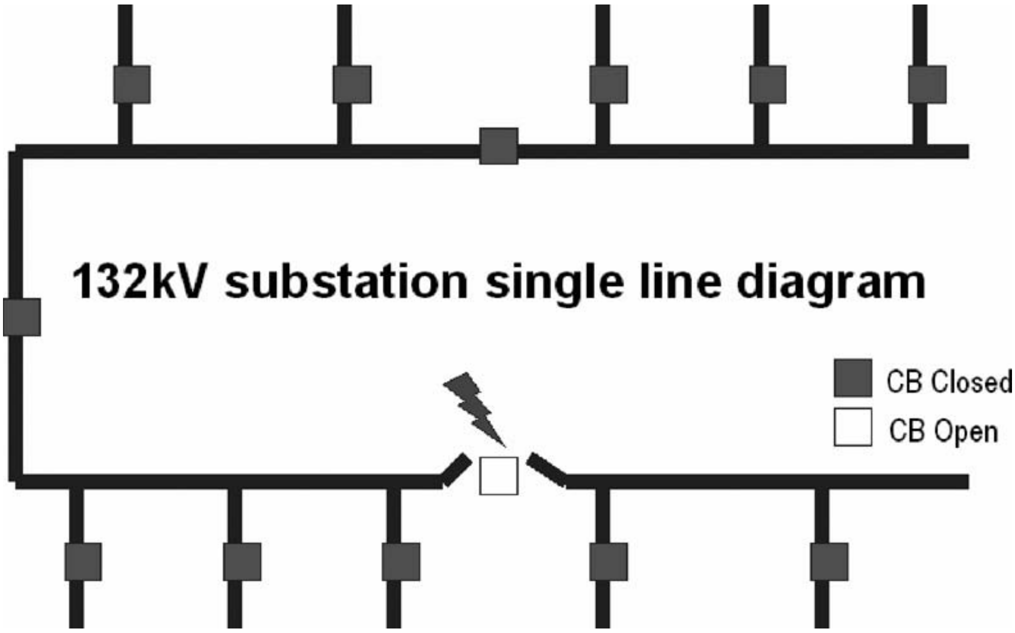


Figure 4. Substation single line diagram after protection trips

A combination of height voltage plant rating issues and network constraints resulted immediately the event occurred. The NCC operator immediately called the AMT and requested assistance and information on what had occurred.

The AMT immediately called the local Maintenance Service Provider and requested immediate site attendance due to the seriousness of the event and the probable need for site attendance.

Within 10 minutes of the event occurring, remote connection to the substation was established and analysis of the substation control system event log was being undertaken.

Within 15 minutes of the event occurring, it had been established that all duplicated bus protection systems had operated on the out of service busbars. Remote access to the bus zone protection relays was established and analysis of the protection disturbance records commenced.

Within 20 minutes of the event, information was provided to the NCC operator that there had been a genuine high voltage fault internal to the bus section circuit breaker, within the overlap of the bus zone protection schemes. The faulted circuit breaker was part of a hybrid switchgear module which included remotely operable isolators on each side of the circuit breaker. A recommendation to remotely open these isolators was provided to the NCC operator to disconnect the faulted plant from the system. While this was being performed, as per Powerlink policy for latched protection systems, consultation with the Asset Manager occurred to brief him of the situation and seek approval to restore the out of service busbars and deloaded feeders.

Within 30 minutes of the event occurring, the NCC operator had remotely opened the isolators to disconnect the faulted circuit breaker from the system. The AMT had remotely reset all the latched bus zone protection relays and advised the NCC operator of Asset Manager approval to restore the out of service busbars and feeders.

Within 45 minutes of the event occurring, the busbars had been restored and feeders reloaded relieving the plant loading issues and system constraints.

At approximately 55 minutes after the event occurred, the local maintenance service provider arrived on site.

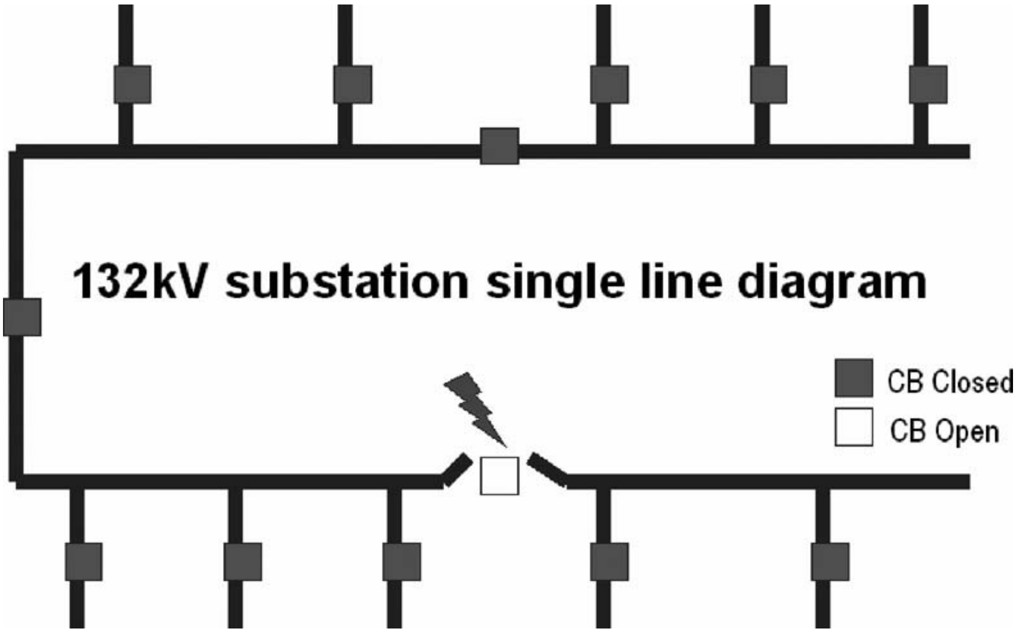


Figure 5. Substation single line diagram after system restoration actions completed.

8.2.1 Case Study 2 Summary

Remote access to the data captured by the secondary systems equipment in the substation allowed the cause of the event to be diagnosed and systems restored as much as possible within 45minutes of the event occurring. The network constraints, and plant loading issues were resolved within 45 minutes of the event occurring.

It is estimated that under the previous fault management strategy the network constraint, and plant loading issues would have continued for at least another two hours until personnel skilled at extracting and analysing data from the intelligent devices attended site and completed these tasks.

9 SUMMARY

Powerlink's Asset Monitoring strategy exploits the advances made in equipment monitoring and communications technology that now allows remote access to intelligent devices in the field from a remote location. Data from these intelligent devices was previously only available by sending suitably trained staff to attend the site, which incurred significant delays in the retrieval and analysis of fault data.

Success of the strategy is not only due to the technology, but the way in which Powerlink has modified its fault management processes to exploit it. The processes are often overlooked in the drive to install new technology. However, unless the processes are modified, the benefits will not be fully obtained.

Powerlink Queensland formed the Asset Monitoring Team (AMT) to implement the data to information framework required for realising the system strategy. The AMT manages the collection, analysis and reporting of information to the business and plant strategy decision-makers. The AMT exists as a small group of engineering specialists, dedicated to the retrieval and analysis of operational data following notification of a network fault. Benefits already being achieved include reducing outage times, exposure to potential liability, fault rectification costs, and, improvements in network reliability and quality of supply.

The new fault management process has been made possible by Powerlink's investment in intelligent substation technologies and communications infrastructure. By accessing substation data from the office, AMT staff are informed about the status of the fault and the health of the plant. Armed with this information, decisions regarding the fault priority, resources required for rectification, and the final decision to return plant to service can be made. These judgments can be made in a short timeframe, in many cases avoiding the need to send personnel to site or before staff arrive at site. It should be stressed however that although the technology is an enabler it is not a solution in itself, as it requires supporting processes as outlined in this paper.

10 REFERENCES

[1] Cukalevski N., Allamby S. et al, "The impact of deregulation and market trading on the requirements for future development of operator training simulators", CIGRE Paper 39-104, Paris 2002.

[2] Australian Energy Regulator, "Service Standard Guidelines 2003", http://www.aer.gov.au/content/index.phtml/itemId/660260/fromItemId/660252

[3] Gabb J., Bell S., "An Integrated Business Approach to Maximise Asset Performance", D2001 paper, Category: Asset Management, Brisbane 2001.

[4] Gabb J., Jones R., "An Innovative Strategy for High Voltage Network Fault Management" D2003 paper, Category Reliability and Quality of Supply, Adelaide 2003.

[5] Bartlett S., "Powerlink Uplinks", Transmission & Distribution World, May 2002.

[6] Borowski L., Moor B., "Modular substation approach is working well at Powerlink Queensland", Transmission & Distribution World, November 2005.

REGULATING ASSET MANAGEMENT THROUGH SERVICEABILITY AND A COMMON FRAMEWORK FOR INVESTMENT PLANNING

Derek Parsons a

[a] Water Services Regulation Authority (Ofwat) 7 Hill Street Birmingham UK B5 4UA

Abstract: Regulation by output (albeit with an associated investment level assumption related to price-cap) rather than input is one of the ways in which the Water Services Regulation Authority (Ofwat), the economic regulator of the water industry in England and Wales, incentivises companies to effective asset management. The paper discusses how such arms-length regulation operates, and considers the appropriate output indicators for such a regime. The paper draws on experience of some ten years of operation and annual monitoring of trends in serviceability indicators, which are the foundation upon which the regulator judges the appropriate financing for capital maintenance at periodic reviews. It describes the UK water industry's risk based approach to capital maintenance planning common framework (CMPCF) jointly developed by companies and regulators in 2002, and Ofwat's approach to its assessment of resulting company business plans. The paper highlights issues arising from the periodic review of prices in 2004 (PR04) and developments needed for the next one in 2009. Companies will need to develop their asset information and analytical tools to enable better forecasting of asset deterioration and the likely impact on service in both medium and long term, to develop a more robust, service driven, risk based approach that delivers economic capital maintenance plans.

Key Words: Ofwat, Water Services Regulation Authority, economic, regulator, serviceability, serviceability to customers, service risk, environment, indicators, trend, capital maintenance, serviceability assessments, service delivered, asset performance, common framework, CMPCF, UKWIR, planning, deterioration modelling, PAS55.

SUMMARY

The aim of this paper has been to broaden understanding of Ofwat's approach to regulating capital maintenance through monitoring of serviceability outputs and its approach to capital maintenance assumptions used for price setting. Inevitably there are references to key documents that mark the sea change in the planning of capital maintenance in the England and Wales water industry since the periodic review of prices in 1999 (PR99) which provide an essential platform for discussion.

Companies not only have to maintain a level of service to their customers and the environment, but demonstrate that service capability of the asset systems is not being compromised for the future. This could happen if performance of key asset groups were to slip through lack of timely investment. Yet it is possible to be over cautious, or inefficiently apply capital maintenance which would elevate prices. So the drive is to demonstrate optimum interventions to maintain the flow of service to customers. The paper tracks recent developments in the regulation of capital maintenance of water companies in England and Wales, (Scotland and Northern Ireland have their own regulation authorities). It sets out Ofwat's approach to annual monitoring and its assessment of business plans at price reviews. The conclusion is that whilst the outcomes of price reviews have been generally satisfactory, there is a need for more transparent, robust deterioration models, coupled with linkage to service delivery and costs to better inform the next price review in 2009 (PR09).

1 REGULATION OF THE WATER INDUSTRY IN ENGLAND AND WALES

Privatisation of water authorities in England and Wales occurred in November 1989, and in setting initial price limits for customers, the UK Government made some broad assumptions about the level of expenditure needed to maintain, or improve and maintain the asset base. At the same time, an independent regulator (in legal terms the Director General of Water Services, but in practice supported by an 'Office of Water Services' or 'Ofwat' was created. His role is to regulate the industry from an economic standpoint to ensure value for money and review and set new price limits at appropriate intervals. In the event, the first price review was in 1994 (PR94) and the two subsequent reviews have occurred at five yearly intervals. Some mergers have occurred, and there are currently 22 companies (plus one with only about 2000 customers) whose prices are set by Ofwat.

Sister quality regulators, now the Environment Agency and the Drinking Water Inspectorate were set up and tasked to set and ensure compliance with standards, and advise Ministers on investment requirements to meet new objectives, mostly emanating from EU directives on potable water and the aquatic environment.

Recent legislation created a corporate structure for water regulation to replace the Director General of Water Services with the Water Services Regulation Authority with effect from April 2006 along with a new duty to contribute to sustainable development. The 'Ofwat' name has been retained for the new organisation.

2 REGULATION OF CAPITAL MAINTENANCE

Ofwat has elected to set price caps for companies every 5 years. Companies are required to submit to Ofwat their business plans for scrutiny which, although focussing on the next five years have a planning horizon for of 15 years (more in some areas). So every five years there is an opportunity for companies to register a need for a change to the planned investment which is then reflected in a new price cap. The price cap is a package encompassing all categories of expenditure and obligations. (There are mechanisms for interim price adjustments, but it is not necessary to digress into this area.)

The main challenges are to:

- monitor and challenge the effectiveness of capital maintenance;
- determine whether there should be a change to past levels of capital maintenance, up or down and by how much; and
- challenge efficiency.

This paper focuses on engineering aspects of optimum capital maintenance assessment. The technical detail of how Ofwat makes the efficiency challenge is a large subject and is beyond the scope of this particular paper, save providing an outline.

2.1 Monitoring and challenging effectiveness of capital maintenance

For a water utility, the principal requirement for capital maintenance is that it is sufficient to maintain service to customers. Ofwat operates arms-length regulation. It does not seek to micro-manage companies because this is resource hungry and ought to be unnecessary in a very mature industry. The regulatory focus is on service to customers. Ofwat, acting on behalf of customers needs to be assured that service level is stable, and is likely to remain so in the future.

Ofwat has developed a set of metrics to monitor company asset systems, and between price reviews it monitors their stability or otherwise through annual review of detailed data returns. These are known as serviceability assessments, which are made at sub-service level (above and below ground asset systems distinguish sub-services). Where a sub-service is assessed as less than stable the company is called to account and is required to set down its proposals to remedy the situation.

The key point here is that at periodic review, the company, in accepting the price cap has also accepted the requirement to maintain (or achieve and maintain) stable serviceability. If the company does this through a set of investments that cost less than Ofwat has assumed, it can retain the financial savings as part of its efficiencies. On the other hand, Ofwat expects a company to invest to maintain stable serviceability, even if this may require more expenditure than assumed at price setting. Thus the specific price limit accepted by each company contains both a profit challenge and downside risk.

The mechanisms to resolve issues are beyond the scope of this paper, suffice it to say that:

- price reviews provide for a timely 'course correction' in funding assumptions (including penalty for failing to deliver, or 'shortfall');
- between price reviews the companies have a financial incentive to maintain stable serviceability efficiently; and
- a company's prestige can be threatened by failure, because Ofwat publishes its annual serviceability assessments.

2.2 Determination of changes to historic levels of capital maintenance

At the 2004 periodic review Ofwat adopted a four stage approach to challenge each company's business plan:

Stage A: Adjustment needed to maintain stable serviceability from a historical perspective.

Stage B: Is the future different?

Stage C: Scope for efficiency

Stage D: Take account of overlaps between capital maintenance and enhancement programmes.

In 2004, the English and Welsh companies 'asked for' over £9.4 billion capital maintenance for the period 2005-2010. Price limits include £8.4 billion. This paper focuses on Ofwat's challenges at Stages A and B. Stages C and D are subjects in their own right and their approach is only outlined in sections 3.3 and 3.4. The remainder of the paper explains and develops the ideas behind Ofwat's Stage A and B challenges and issues that are emerging for resolving at or by the next periodic review.

Stage A is essentially a long term historic average cost, corrected for price base. Companies were allowed to make adjustments in their business plan submissions for whether or not this was typical and whether serviceability could be maintained (see section 8). Ofwat made its own assessment, and where the company view was greater expenditure than Ofwat's view at Stage A, the difference received further scrutiny by being fed into Ofwat's Stage B analysis (see section 9).

Stage B is essentially Ofwat's way of considering the validity of a change from the past rate of expenditure into the next period(s). In most cases, companies tend to seek more going forward. But how robust is the engineering business case?

2.3 Stage C - Efficiency challenge and incentive mechanism

Ofwat commissioned several studies and reviewed contemporary work of others to come to a view on the **scope for efficiency**. This is the amount of capital expenditure that the water industry in total can be expected to save through capital efficiencies of one sort or another in the period 2005-10. Ofwat uses a carrot and stick approach, whereby the scope for efficiency is divided into two parts. It removes one part on behalf of customers (stick), leaving the other part as an incentive for the company (carrot). The size of the carrot and stick for each company depends on how near it is to the efficiency frontier.

Relative capital expenditure efficiency: this is the (percentage) difference in efficiency that has been observed using historic (and forecast) data between the capital expenditures of all the companies. The company with the lowest measured capital costs (in a specified category or basket of indicators) becomes the 'benchmark' company. Ofwat assesses relative capital expenditure efficiency in two ways:

- using econometric modelling of total capital maintenance expenditures (averaged over a number of years) and various explanatory variables – this is known as the Capital Maintenance Econometric Analysis; and
- measuring the differences in company's capital scheme unit costs (procurement costs) – this is known as the Capital Expenditure Cost Base Analysis.

The company is allowed to keep its out-performance for five years. The details of the econometric models and cost base unit costs are set down in the 2003-04 issue of Ofwat's Water and sewerage unit costs and relative efficiency report. It covers both capital and operating costs, which also use benchmark econometric modelling. The impact at PR04 was to shave some £0.5 billion from capital maintenance assumptions across the English and Welsh companies, which makes it a big subject.

2.4 Stage D –Adjustments for Enhancement Expenditure - Benefit to Serviceability

Stage D is the final adjustment made to derive the required level of capital maintenance expenditure going forward. The Stage D adjustment allows for the impact on future serviceability from planned expenditure on enhancement projects, whilst taking into account any change in proportional allocation of enhancement expenditure to capital maintenance going forward.

3 ECONOMIC CONSTRAINTS TO LIMIT THE PRICE PAID BY CUSTOMERS

Ofwat holds the view that prices, which support maintenance of the vast asset base, should be no higher than they need to be to limit the price paid by customers. In the 1999 Periodic Review of Prices (PR99) the Director General of Water Services judged that the water companies had not moved sufficiently in their business practices towards a more efficient approach to capital maintenance. After setting prices for the period 2000 – 2005 he wrote, in April 2000 an open letter to Managing Directors of the water companies of England and Wales, MD161 [1]. He recognised that companies had put much effort into improving systems of information for asset management, resulting in more effective direction of capital maintenance activity. But the key message was that the companies' economic analyses fell short of expectations. Thus he wrote:

3.1 Extract from letter MD161 to Managing Directors, April 2000

'Each company needs to demonstrate how the flow of services to customers can be maintained at least cost in terms of both capital maintenance and operating expenditure recognising trade off between cost and risk, whilst ensuring compliance with statutory duties. Appraisal of capital maintenance, operating expenditure and risk can be compared using discounted cash flows. All such appraisals would need to be set in the context of the framework of maintaining serviceability to customers. Such an approach should have been used to justify the future levels of capital maintenance included in the business plans.

It would have been helpful to include commentary on the material elements of the economic appraisals undertaken such as the:

- Cost of any potential loss of serviceability to customers, including consideration of risk scenarios and their probabilities as well as illustrations of how serviceability to customers would decline, if the activity was not undertaken;
- Impact on operating costs of capital maintenance activity, before and after assets are renewed;
- Circumstances surrounding the timing of asset replacement;

- Impact of obsolescence and newer lower cost technology; and
- Any terminal values and the discount rates assumed.

This list is not exhaustive. The economic appraisals would also need to be balanced and justified against a strategic top down approach used to assess, for example, the impact of alternative scenarios on company financing, including an analysis of past and prospective accounting charges. Setting out examples of economic analysis in this way would go a long way towards demonstrating why companies have taken the view that the levels of capital maintenance since, say 1980, have been sub-optimal. This approach could also indicate the extent to which levels of capital expenditure should change, up or down, for future capital maintenance to be economic.'

So Ofwat puts the customer, not the engineering, centre stage. Exponents of PAS 55, take note of these stakeholders!

4 MAINTAINING THE FLOW OF SERVICES TO CUSTOMERS: 'SERVICEABILITY'

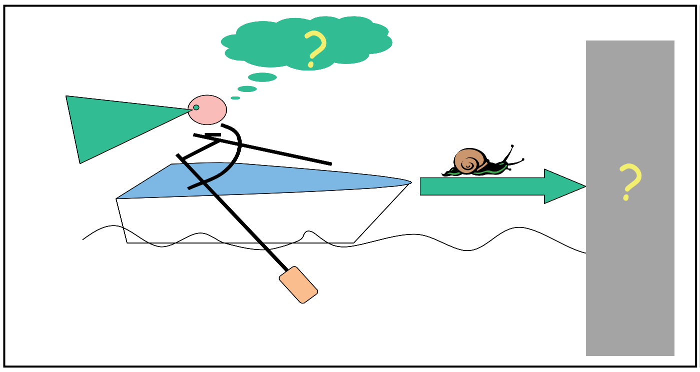
Specifically, the term used by Ofwat is 'Serviceability to customers'. This is NOT the same as serviceability of the assets deployed to deliver service – but there are close relationships between what service the customer wants and the ability of the assets to deliver it. Maintainers of assets sometimes find it is easy to forget why the assets are there in the first place.

The term encapsulates the concept of the continuing capability of a company's asset systems (and the associated operational activity) to maintain the level of service experienced by its customers and, where relevant, the receiving environment. The regulatory focus is on delivery of customer needs (both now and in the future) and not on maintenance of the assets to a specified state for their own sake. The 'right state' should essentially be determined by economics, albeit the default position in the absence of robust economic argument for change should reference to a suitable historic level.

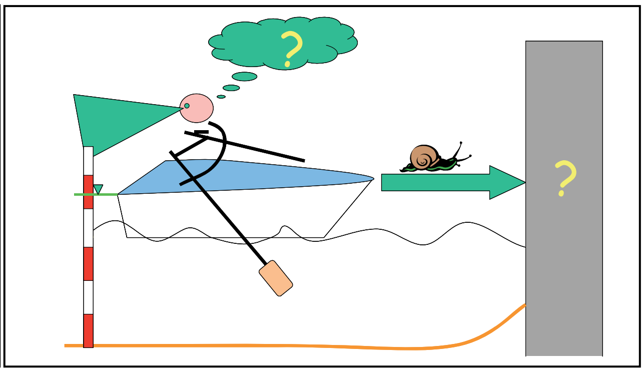
4.1 Assessing serviceability: Anticipating future problems from an historical perspective

Regulatory serviceability indicators have come in for a good deal of adverse criticism in the UK, cited as being 'backward looking'. The author contends that just about everything that can be forecast about the future is based on past experience or past observation, and therefore invites the reader to re-consider whether they are, after all, useful by considering the following thought experiment.

Suppose you are in a lake, rowing a boat, alone. You are facing backwards and have no view of the boat's forward, but slow progress.

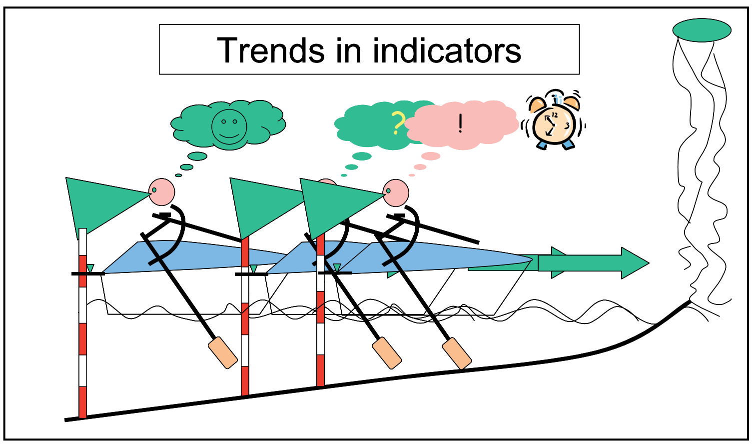


How might you anticipate arriving at the shore?



Perhaps measuring the depth might give you some idea?

At intervals you make a series of depth observations. Perhaps monitoring the trend in the depth provides warning of your arrival at the shore? The conclusion is surely that rearward view of appropriate indicators, monitored at regular time intervals does indeed reveal useful information about change. (The analogy can, of course, be trivialised, because all you do is look over your shoulder. If only looking into the future was that simple!)



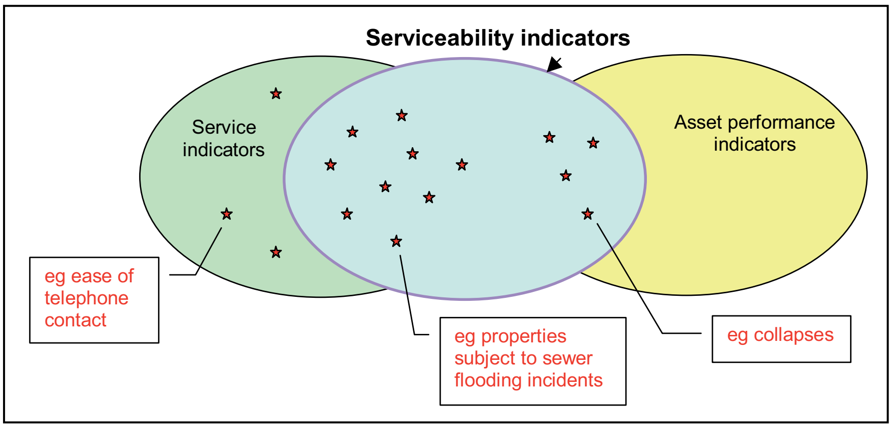
Put in its simplest form, such indicators provide the basis of assessing whether or not there has been change. And, stepping right away from the row-boat analogy into asset management, consider next what the inputs have been to maintain the status quo or cause change to our suite of indicators (a suite because more than one is required, as will be explained later).

Serviceability is considered at company level, by sub-service. The aim is to ensure that the trend in serviceability remains at least stable, or if the start point is less than stable, then to achieve and then maintain stable serviceability.

Note the key word is **trend**. Absolute values of indicators are not important, although it is useful to compare and contrast each company's performance and to build a national picture. The assets being considered have relatively long service lives, and deterioration is relatively slow. Thus annual monitoring provides the opportunity to take timely corrective action.

5 APPROPRIATE SERVICEABILITY INDICATORS

The figure below illustrates from where indicators of serviceability are drawn - courtesy of United Kingdom Water Industry Research (UKWIR) Report 02/RG/05/3 [2]



Clearly, a service indicator such as ease of telephone contact has little direct connection with capital maintenance needs. The basket of serviceability indicators comprises a selection of measures of service delivered and asset performance. A basket of measures is needed due to the complexity of asset systems and interdependencies. A combination of service and asset performance indicators is needed in order to capture current and likely future service capability. Each type as has certain characteristics and limitations:

5.1 Service indicators are influenced by operational activity, and trends are to be assessed with caution. They inform on the capability of asset systems in the shorter term. The expectation is that trends are stable or improving.

5.2 Asset performance indicators inform on capability over the longer term, that is whether the current service delivered is likely to be sustained. The expectation is that, whilst trends are stable or improving, they may deteriorate to a degree that does not compromise economic provision in the long term. In practice, the default trend for asset performance indicators is stable or improving. Asset performance indicators may have a direct or indirect impact on service delivered. They tend to be symptomatic of the state of the assets, but still can be influenced by policy on operational activity, for example increased leakage control activity in the case of water mains burst repairs.

5.3 Examples of serviceability indicators

Service indicators: properties with extended supply interruptions, properties receiving low pressure, properties flooding from sewers, pollution incidents. Compliance, such as the proportion of treatment works that are out of compliance.

Asset performance indicators: burst water mains, sewer collapses, mean time between failure. Sub-threshold indicators that indicate potential underlying problems but where the works has not failed. Examples are events where a threshold value, say half the permissible value has been crossed.

5.4 Issues around serviceability indicators

A very important point is that it must be the same data that is collected, that is, to the same definition, over the period of time which is assessed. This must be watched closely, because both the Levels of Service and financial data which are essential for the historical perspective analysis are collected for other uses, and there is often a desire to 'improve' them for those uses.

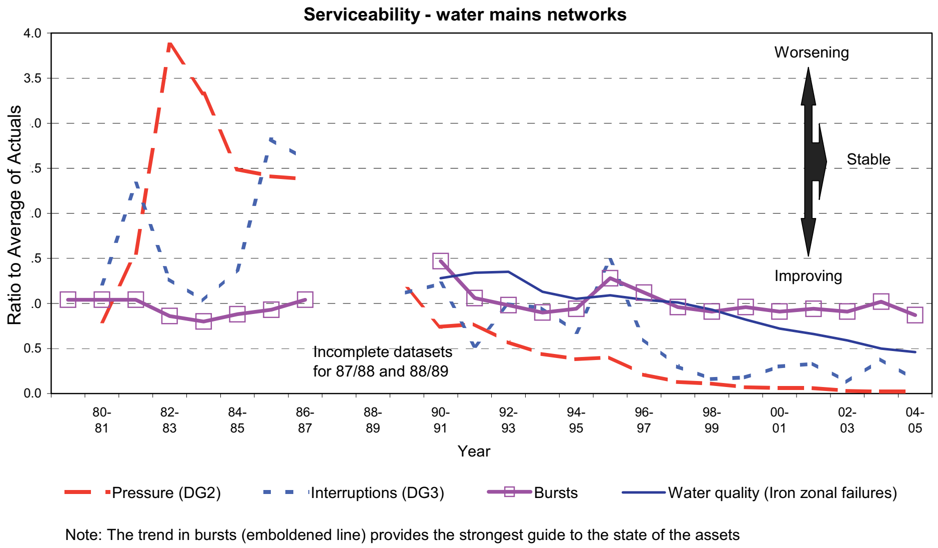
But of course a change in definition invalidates a time trend, and reliable and defensible time trends are the essence of the historical perspective analysis.

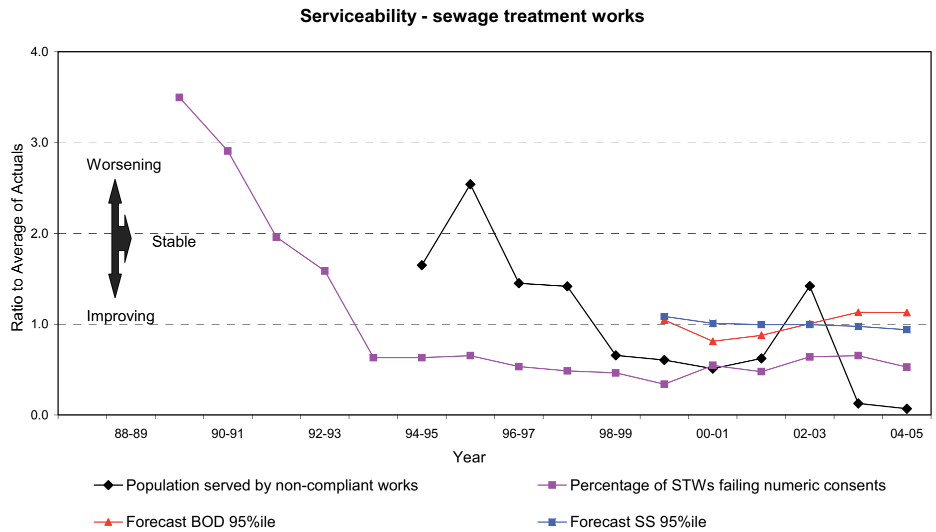
Although there will always be interplay between capital and operational maintenance it is assumed that the deferment of capital maintenance by increased operational maintenance will not be effective over many years. Nevertheless, for surface assets it is wise to look at the pattern of operational maintenance alongside the graph of capital maintenance. It is also important to check companies' explanation of recorded treatment failures. These are generally few in number and experience in the 1999 price review was that companies explained many of them as due to short-term events which had been quickly remedied. In fact they argued that capital maintenance had been adequate even though the serviceability analysis may have indicated otherwise. A conclusion Ofwat was pleased to accept.

From 2003 Ofwat added a limited number of additional indicators that aimed to inform future performance and to widen coverage of asset types [3]. It planned a further review after PR04. The review has now been carried out by UKWIR (2005-06), with Ofwat represented on the steering group. The report is imminent and Ofwat will shortly respond to its findings.

6 ASSESSMENT OF SERVICEABILITY

Ofwat's initial assessments are published annually in August/September in its 'Financial performance and expenditure of the water companies of England and Wales' (FPE) report. Below are examples of Serviceability Graphs at Industry Level for England and Wales, for water infrastructure and sewerage non-infrastructure sub-services respectively.

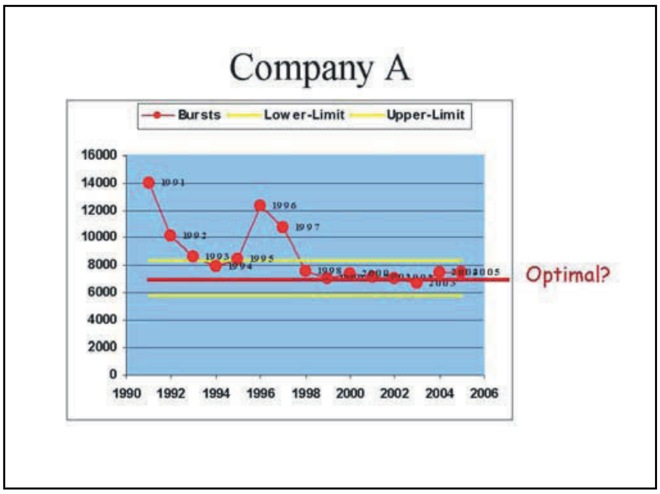




6.1 Assessment marks

The current marking scheme for serviceability trends is, best to worst: Improving, Stable, Marginal, and Deteriorating.

Data is inherently variable, often due to weather effects and care needs to be taken when considering what a normal range of operation is. Because underground assets in particular deteriorate slowly it is not easy to pick out underlying trends. Ofwat uses a suite of statistical tools to help inform its judgement. The statistics toolkit is not a substitute for judgement, simply an aid. It also provides a platform for discussion with companies on their view of the optimal levels for a particular indicator.



6.2 What is meant by 'stable' serviceability?

Serviceability is deemed to be stable when the assessment of trends in a defined set of service and asset performance indicators demonstrates that service is in line with the reference level of service and, by inference, is likely to remain so into the future. Note that there is no implication that the standard of service to customers or asset performance is satisfactory. These are separate issues, which should be addressed through a change in expenditure for improvement or economic argument.

The reference level of service is determined from a specific sub-set of public health, environmental and customer service indicators. Service indicators reflect the degree of compliance with statutory regulations, regulatory and company standards and customer preferences.

Asset performance indicators, measured at system level, are drawn from a specific sub-set of measures that inform current and future levels of service. Stable serviceability normally requires that asset performance is in line with the reference level of asset performance. Unless demonstrably sub-optimal or atypical, the reference levels of service and asset performance are taken as the best historic levels achieved by the company.

6.3 Issues around achieving and maintaining stable serviceability

As the reader may imagine, things begin to get interesting for regulator and regulated when a company begins to fail to deliver stable serviceability. The serviceability assessments enable Ofwat to provide feed back to companies on how it thinks they are doing. Annual monitoring is normally sufficient for Ofwat's purposes, and, despite the indicators being high level, the process has, on occasion uncovered underlying shortfalls in asset management practice. It has thus become a powerful tool.

The assessments enable a regulatory response where companies are beginning to fail, and can provide evidential support for financial penalty or other legal sanction in the most serious of cases of shortfall. This provides motivation for companies to intervene with more effective monitoring and intervention (capital or operational expenditure), and yet still allows the company to decide what to do.

In section 3 it was explained that the periodic review of prices provides scope for a 'course correction' if it is deemed that funding from customers should be changed to facilitate an economic level of capital maintenance. And once prices have been determined, and the company accepts (if not it can appeal to the Competition Commission for a new determination), it is up to the company to deliver stable serviceability. This is irrespective of cost to the company (subject to a high ceiling), since the company accepts this as a risk. But note that the 'course correction' at periodic review does not necessarily cover the cost of recovery to stable serviceability where there has been a shortfall in previous periods. The change is primarily to reflect what funding is needed by an efficient company to maintain stable serviceability going forward.

The risk to serviceability accepted by the company at price review is to some extent a shared one. This is because there is incomplete knowledge, for example in the behaviour of assets and in the economic level of capital maintenance, and serviceability might in any case be changed by events that might render a rapid return to previous service levels impracticable.

In the past, companies have not been overtly penalised, for example where they have spent above Final Determination assumptions and still failed to maintain stable serviceability. Companies are expected to improve their knowledge of their asset base, for example through an improved application of risk based principles espoused in MD161 in the period to 2010. This should lead to better assessments of need by the companies; and Ofwat will be less tolerant of serviceability shortfall.

Where a company is beginning to fail to deliver on serviceability, Ofwat deploys a staged approach to effect corrective action by the company, including in appropriate cases a requirement for setting down action plans with milestone dates for activities aimed at achieving and maintaining stable serviceability. Both the historic level and new information may inform the degree of recovery of serviceability that is required following a period of deterioration. And in the absence of robust justification to the contrary, the historic levels provide the reference level for serviceability.

This low level but firm regulatory action is aimed at bringing defaulting companies to heel and to prevent a situation developing that could cause a general or catastrophic failure of service.

7 CAPITAL MAINTENANCE PLANNING – A COMMON FRAMEWORK (CMPCF)

In 2002 the UK water industry, through UKWIR, responded to MD161 with the development of the capital maintenance planning common framework (CMPCF)[2]. Ofwat regarded the CMPCF as a large step in the right direction, and redrafted its reporting requirements for PR04 business plan submissions to align with it.

The common framework is essentially a three-step process (similar but distinct from Ofwat's 4-Stage A, B, C, D business plan assessment):

A. A Historical assessment, comprising an expenditure and service & asset performance review.

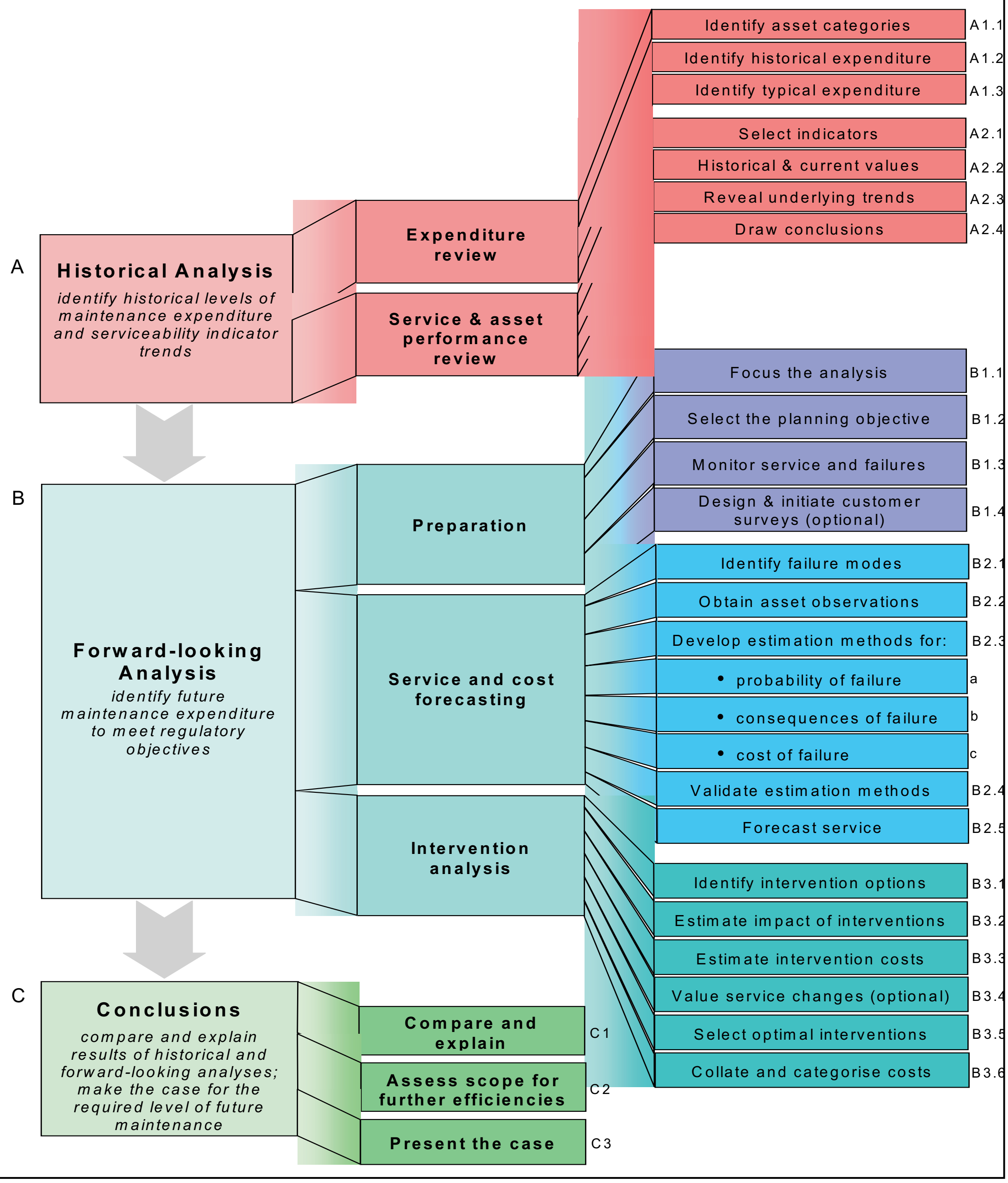
B. B Forward look, considering service risks and cost forecasting to develop optimal interventions.

C. C Comparison of the two approaches, drawing conclusions, look for further efficiencies and make the case for the required level of capital maintenance.

Application of the CMPCF includes two alternative planning objectives – a cost-effective approach or a cost-benefit approach. The cost–effective objective aims to provide steady or improving service to customers at minimum cost to the water company. The cost-benefit objective aims to provide the level of service to customers that represents an economic balance between the value to customers of the service provided (perhaps in terms of willingness to pay) and the associated costs to the water company (and thence its customers).

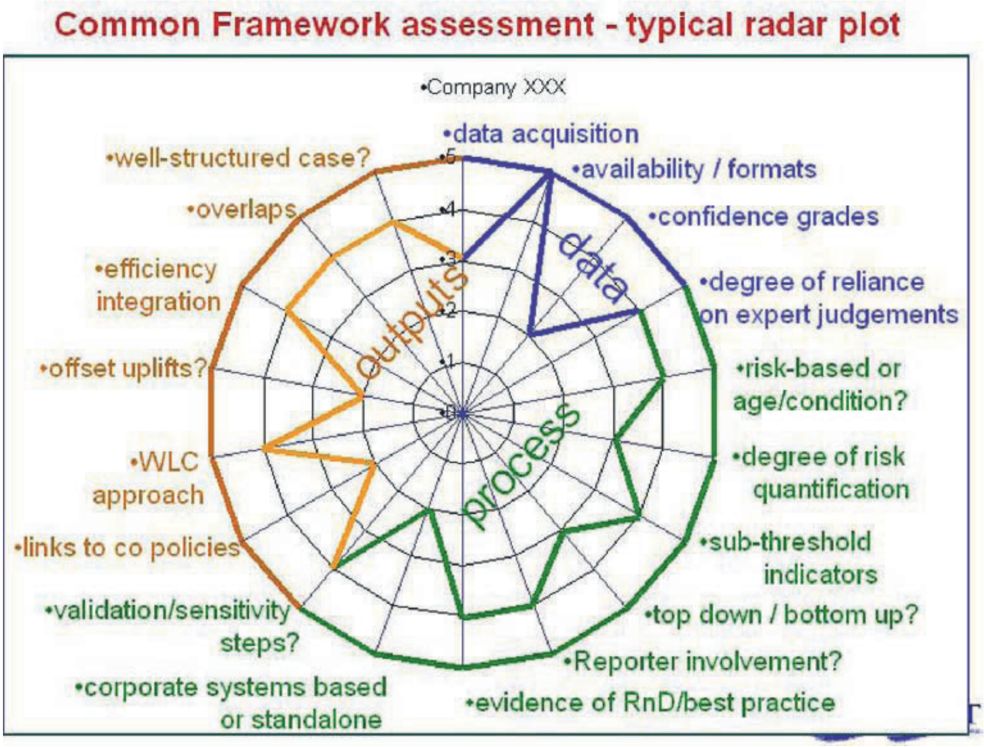
In fact there are over twenty steps at the detailed level, as the figure below shows. The main innovation is in the second step, 'B' that uses the analysis of risk (that is the probability and consequences of asset failure) to inform the economic analysis. Previous practice in UK water industry has been to attempt to convert condition and performance grades (captured with the asset inventory) into remaining asset lives and hence capital maintenance needs. The problem with that approach is that the grading approach tends to be over subjective, generalised and asset centric. Under the CMPCF companies must look to identify specific failure modes by asset type, and at an appropriate level of granularity, obtain the relevant asset observations, including historic failures, consequences and costs, then develop robust estimation methods for probability of failure and consequences for customers. These, together with service valuation studies for the cost-benefit approach, are developed into service and cost scenarios to inform the optimisation of interventions. By evaluating in this manner, expressed in economic-risk measures, the analysis is better focussed. The best exponents can locate where investment will be most effective and may build the programme of works at the time of drawing up the business plan. Unfortunately, in the short term sufficient data is not always available and judgement is needed (nor will judgement be entirely eliminated) in place of a history of asset observations. The process gives a powerful framework for the company to identify what asset observations it needs to improve its forecasting to facilitate optimum interventions.

Capital Maintenance Planning Common Framework (CMPCF)[2]



8 ASSESSMENT OF THE APPLICATION OF CMPCF PRINCIPLES

Ofwat's four stage approach provides a means of evaluating the application of a company's application of CMPCF principles. To challenge stage B uplift over past expenditure levels Ofwat assessed each company's final business plan against 18 criteria. The criteria reflected the aspirations of MD161 and the agreed CMPCF. The criteria cover the three broad areas: data quality, forward-looking analysis process adopted and the approach to outputs. The criteria assessment became known as the 'Common Framework Assessment'. As with serviceability, this was done by sub-service. There are 64 sub-services in total among the companies – 10 water & sewerage co.'s x 4 (water/sewerage above & below ground) plus 12 water only co.'s x 2 (above & below ground).



The criteria were weighted by pair-wise comparison of relative importance. Sensitivity analysis showed that total scores were not overly sensitive to the weights. For each sub-service, the 'score' was allocated into one of five ranking bands ('Leading' to 'Trailing'). The bands determined the factor applied to the uplift in Stage B expenditure (post screening to exclude exceptional items assessed on their own merits). The main problem for companies at PR04 was their inability to find suitable data, leading to more 'expert judgement' than desirable, and the inability to adequately identify operating costs at plant level to be able to get at whole life cost of ownership.

Band	Description	Score range	Factor	Number of sub-services
A	Leading	>= 70%	100%	5
B	Above intermediate	60 – 69%	75%	22
C	Intermediate	50 – 59%	50%	29
D	Below intermediate	40 – 49%	25%	7
E	Trailing	< 40%	0%	1

A 'leading' company will understand the risk profile of its assets and be able to assess the investment needed to manage this risk in the future. A company assessed as 'trailing' will have failed to demonstrate a robust risk-based approach for future investment in this sub-service. Companies with scores close to band limits were reviewed to confirm their position.

At the periodic review of 2004, this challenge on the quality of business plans was limited to the companies' forward look, that is (broadly) what companies wanted in addition to what they'd had in the past. And even then after exceptional items were excluded. So only around 20% of company's capital maintenance expenditure was subject to the test.

8.1 Issues around CMPCF

The impact of Ofwat's challenge at stages A and B amounted to some £0.5 billion, mostly at stage B. UKWIR has reviewed the CMPCF application [4] and outcomes post PR04, as has Ofwat. As always, Ofwat consults with the industry [5] and, at the time of writing this paper, is doing so again [6]. The Ofwat view is that there should be a stronger challenge to historic expenditure, to develop a fairer playing field. Options include simply improving understanding of reported historic expenditure, econometric benchmarking to cap Stage A, assessing part or all of historic expenditure in the Stage B challenge, analysing asset stocks and condition and age profiles. Econometric benchmarking has implications for the efficiency challenge and would need to be considered in conjunction with revisiting stages C and D.

The challenge for Ofwat is to encourage appropriate behaviours in business management; to challenge old habits and drive efficiency. The evolving approach gives companies strong incentives to use sound asset management processes. Ofwat will do this by strengthening the incentive through its assessment processes. Ofwat also believes that the CMPCF approach, whilst it has been developed for capital maintenance could potentially inform a more general application to all capital investment drivers. This is because of its focus on service to customers and risk in order to plan the scale and scope of investment.

8.2 CMPF and PAS55

Both CMPCF and PAS55 are risk-based approaches to maintenance. In the view of the author the key difference between PAS55 and CMPCF is that PAS55 is asset centric, whereas CMPCF is overtly centred on service to customers. In the water industry this is particularly important, because there are so many facets to the service arising from the need to maintain product quality as well as continuity of service.

Ofwat uses its assessment of how well companies applied CMPCF principles its determination of price cap in respect of capital maintenance. Ofgem, the UK energy regulator, in contrast, has used PAS55 for re-assurance and not directly to influence its price review judgements. PAS55 is seen by Ofgem as good practice to aspire to. Interestingly, Ofwat's idea to use radar plots to compare current business practice was 'borrowed with pride' from Ofgem. Ofgem prefers to use other tools for price setting, due to a quite different regulatory environment. For example, competition exists on load (demand) in the electricity sector which strongly influences price. Therefore, for the moment at least, there seems not to be an incentive for Ofgem to further motivate electricity and gas companies develop their application of the principles in PAS55. It will be interesting to monitor developments.

9 THE FORWARD LOOK – GAINING CONFIDENCE IN DETERIORATION MODELLING

At PR04, Ofwat's four stage approach provided for both a historical review (Stage A) and a forward look (Stage B). The forward look embraced the CMPCF, and encouraged companies to demonstrate what is different from the past, going forward. Whatever approach is taken, some form of forecasting of asset behaviour is needed, in order to distinguish either a change in the rate of deterioration from no change, or a change in the volume of assets needing renewal. It is also essential to link asset behaviour (and failure modes) to service delivered. Some software tools were developed of varying complexity for company business plans at PR04. But all such tools are vulnerable to the axiom 'rubbish in equals rubbish out'. The drive now is not so much to develop new models, though the current initiatives through UKWIR are welcome, particularly on deterioration of sewers, but rather to gain confidence in deterioration modelling and application as a process. The need is to:

- expose the assumptions in the models to rigorous scrutiny;
- broaden the understanding of models and modelling techniques;
- peer review the models;
- define criteria for good data for the particular model;
- do a gap analysis and look for solutions to fill them;
- carry out sensitivity analyses on model outputs;
- validate the models by recognised techniques; and
- demonstrate the application of the models
- link 'failure' to service failure and customer preference

Whilst there is a degree of commercial interest that constrains this approach, the models earn their keep best when Ofwat has been convinced of their reliability, and, paradoxically their limitations. At the end of the day, judgements still have to be made, since, like the man in the rowing boat in the thought experiment he cannot see into the future. There is always an element of faith.

10 CONCLUSIONS: ARE THE OUTCOMES OF PR04 SUSTAINABLE?

It is evident from successive periodic reviews of prices that, on the whole, companies have put forward in their business cases more expenditure than was, in the event, actually needed to maintain stable serviceability. But it does not take a genius to wonder, in the case of infrastructure whether the amount assumed is sustainable. The simple manipulation of figures - annual investment divided into the asset value gives an alarming view that the infrastructure assets will need to last for several hundred years. How is this right? Is it right?

The quick answer is, for the moment probably yes, based on the experience of today. 'It ain't so broke, so why fix it?' And, due to the well founded belief that infrastructure assets tend to deteriorate slowly, the capital maintenance assumptions can be ramped up over successive price reviews if deterioration in serviceability is detected. Also, some of these assets will outlive their usefulness due to redevelopment etc, and therefore, there is no need to think that every pipe is going to need replacing anyway.

The more measured answer is that the companies are managing the asset base quite well in an increasingly proactive way between price reviews. There are concerns, particularly for those companies with less than stable serviceability, which they are required to correct by 2010. These are being monitored closely. There is a more general need for companies to get a better understanding of what any ramping up of future expenditure might need to be, and whether or not the problem will remain manageable without earlier pre-emptive investment. It is this nagging uncertainty that drives the need to develop deterioration models for an aging (but generally adequate) underground infrastructure. And there are similar uncertainties surrounding the renewal requirements from the new above ground/non-infrastructure assets constructed since privatisation. These have very different characteristics from the earlier assets that they either replaced, or were built as new/advanced processes to meet higher standards. Thus for these assets, historic expenditure may not be the best guide going forward, a point well recognised by the introduction of the CMPCF. Companies will need to develop their asset information and analytical tools to enable better forecasting of asset deterioration and the likely impact on service in both medium and long term, to develop a more robust, service driven, risk based approach that delivers economic capital maintenance plans.

All assets should be there for a purpose, and do not need to be maintained for their own sake. So whatever modelling of deterioration of assets is done, there has to be recognition that the risk to service from asset system failure is the prime concern. So the need for PR09 is to show how models for asset deterioration link to service consequences of failure and associated costs (remedies and impacts) in order to demonstrate that the company has identified robust optimal restorative interventions.

Ofwat works with and encourages the industry to develop its approach. The Ofwat website: www.ofwat.gov.uk provides a wealth of information through its publications and public consultation processes, much of which is referenced in this paper.

References/Bibliography

[1] Letter from Ofwat to Managing Directors MD161 12 April 2000 MAINTAINING SERVICEABILITY TO CUSTOMERS

[2] UKWIR report ‘Capital Maintenance Planning a Common Framework’. Ref No. 02/RG/05/3 UK Water Industry Research Limited. 2002. ISBN 1 84057 265 5.

[3] Letter from Ofwat to Regulatory Directors RD15/02 8 May 2002. AN UPDATE ON SERVICEABILITY INDICATORS AND MEASURES

[4] UKWIR report ‘Capital Maintenance Planning Common Framework: Review of Current Practice’ Ref. No. 05/RG/05/14

[5] Letter from Ofwat to Managing Directors MD212 2 February 2006 ASSET MANAGEMENT PLANNING TO MAINTAIN SERVICEABILITY

[6] Letter from Ofwat to Regulatory Directors RD04/06 27 March 2006 DEVELOPING OUR PROCESS FOR ASSESSING CAPITAL MAINTENANCE REQUIREMENTS

Acknowledgements

Dr Colin Sinnott, Consultant. Inventor of the historical perspective of serviceability and capital maintenance.

UK Consultants who have contributed significantly to the development of approach from circa 2000:

WS Atkins; Ewan Group; Mott MacDonald; SMC, Stone & Webster; Tynemarch Systems; WRc/BHr.

UKWIR for facilitating development of the CMPCF, and many subsequent projects on-going to assist the development of capital maintenance planning.

The Water Companies of England and Wales, sister Regulators Drinking Water Inspectorate and the Environment Agency, for providing data and technical input into the development of serviceability indicators and the CMPCF.

Colleagues at Ofwat for encouragement, peer review and support in writing this paper.

DELIVERING RELIABILITY IMPROVEMENT & EFFICIENCY THROUGH EMERGING DIAGNOSTIC TECHNIQUES AT POWERCOR.

Kieran Skelton

Asset Manager, Powercor Australia.

Abstract: This paper outlines some of the benefits achieved at Powercor Australia through the evolution of asset management methodologies and their application. It considers recent history and the knowledge gained throughout the journey with particular emphasis upon efficient work packaging through improving diagnostic techniques.

Key Words: Asset Management, Diagnostic Techniques, Methodologies.

1 INTRODUCTION

Powercor has an electrical network valued in excess of $2.2B and supplies electricity to over 620,000 consumers across Western Victoria in the South East corner of Australia.

In 2001 Powercor reviewed its Asset Management Strategy objectives to bring a greater emphasis onto the need for a rapid reliability improvement in the operating of the electrical assets. This improved reliability was to take the form of a reduction in the planned and unplanned outages (breakdowns) that impacted upon the assets.

Preceding this review the organisation was based a process related structure that focused largely upon the standardisation of processes and tasks undertaken. Specific benefits that were targeted related to short term financial gains and prudent capital investment. The objectives during this phase were to largely deliver a reduction in costs to consumers following disaggregation and privatisation of the industry in Victoria.

From 2001 until 2004 the revised Asset Management Strategy Objective was to deliver a step change in reliability performance with a relatively minimal step change in capital investment. This is a challenging objective that is compounded in terms of difficulty when attempted in a capital intensive industry such as electrical distribution.

This paper describes how the benefits of emerging diagnostic techniques such as corona detection, and thermography can be used as inputs to a structured Asset Management framework to improve reliability performance and efficiency of resource and capital on a large dispersed asset such as an electrical distribution network.

It also considers some of the next generation of benefits that are being targeted, and the process to move towards best practice, over the next five year period.

2 COMPANY PROFILE AND VISION

Powercors vision is “To be a leader in the distribution business in Australia with outstanding performance. We strive to excel in financial performance, productivity, supply reliability, customer service and community perception”. This translates to the delivery of increasing benefits to shareholders, customers and the communities we serve until we are leaders in each category.

Powercor Australia is the largest of the five Victorian electricity distribution networks in terms of geographical service area and customer numbers. Powercor services some 620,000 customers over an area of 150,000 km2 stretching from the Western suburbs of Melbourne to the South Australian and New South Wales borders. It has assets of $2.2B and is jointly owned by the Cheung Kong group from Hong Kong and Spark Infrastructure Limited a publicly listed company in Australia.

The assets largely comprise of 512,000 poles, 76,000 km of overhead lines, 6,000 km of underground cables, 74,000 transformers and 63 zone substations which are used for electricity distribution.

Figure 1: Powercor Distribution Network Coverage

With the move to a privatised and deregulated electricity industry in Victoria the need to consider both commercial and customer service needs as a measure of a company performance are key drivers for the electricity companies. Since 1995 the industry has progressively been separated into a market driven retail and generation component as well as a regulated transmission and distribution component of which Powercor is a part of the latter.

3 REGULATORY ENVIRONMENT

Companies owning a regulated distribution system, or electrical network as it is often called, are required to obtain and operate within a distribution licence which is linked to various operating and performance standards. These standards are legally binding and legislated within the Victorian electrical distribution code. The Essential Services Commission (Government Regulator) has a charter to ensure compliance occurs by all distributors and oversees the distribution code. The regulator reviews the code every five years to reset the allowable distribution tariffs, and customer service levels, electrical distributors are required to meet in the next review period. The 2001 to 2005 review was completed in October 2000 and the amendments adopted by the regulator required a step improvement in the levels of reliability presently being delivered followed by a level of incremental improvement equating to 20% over the next 5 years as shown in figure 3. In addition to these reliability improvements there was to be a tariff reduction of approximately 20% directly impacting upon revenue.

4 POWERCOR ASSET MANAGEMENT STRATEGY EVOLUTION

The original Asset Management Strategy (AMS) and model already in place at Powercor had been developed with Price Waterhouse Urwick's assistance as an implementation partner in 1997. Underpinning the strategy at that time were 63 Asset Maintenance Policies that had been developed by using the Reliability Centred Maintenance II (RCM II) methodology which was very successful in contributing to Powercor's reduced maintenance and asset inspection costs over the following years.

Powercor had implemented SAP R/3 integrated business software utilising the SAP Plant Maintenance module. The Plant Maintenance module in SAP performs a Works and Asset Management function and was a key factor in the successful implementation of the original Asset Management Strategy. With the implementation of SAP R/3 a process structure was used as the organisational hierarchy to deliver a standardisation of tasks and processes.

The Asset Management model as shown in figure 2 was designed to look at assets and determine how performance improvement through a whole of life approach can be maximised. This goal is reflected in the Asset Management "mission statement" and objectives as follows.

"Asset Management is the structured and systematic approach to acquiring, managing, maintaining and disposing of assets so that Powercor's customers' and business needs are met, and its obligations to share holders, the community, government and regulator are satisfied.

Asset management will address:

- *reducing the need for new assets by adoption of "non asset" solutions where it is cost effective ;*
- *ensuring that existing assets are effectively and efficiently monitored, maintained and utilised;*
- *optimising total life cycle cost of assets from acquisition operation, and maintenance, to disposal; and*
- *establishing clear accountability and responsibility for assets and their performance."*

Between 1997 and 2000 the AMS had delivered a financial benefit and contributed to a reliability improvement trend that was positive although incremental at best. Similarly the process structure in place had delivered significant benefits in the standardisation of processes and bringing consistent and common levels of reporting and data acquisition into the company that previously did not exist. It was appropriate for the regulatory targets and shareholder objectives that were set at that time.

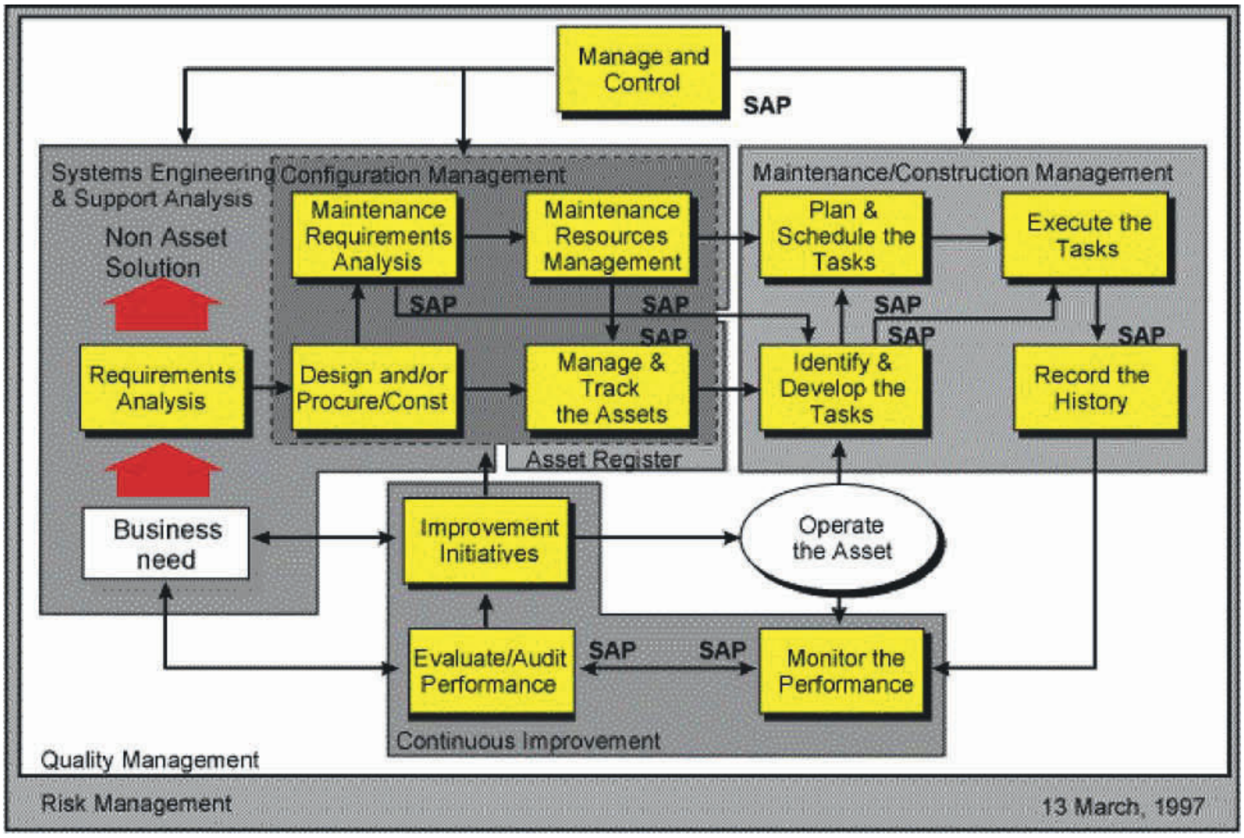


Figure 2:Asset Management Model

In 2001 with new owners and a new regulatory period came new challenges. The new shareholder influence and vision required a step forward in reliability performance with very prompt and financially efficient implementation path.

The Asset Management model was seen as a sound platform having served its purpose well over the previous 3 years and was retained. The mission statement and objectives were reviewed to reflect the new regulatory period and increased emphasis on reliability performance. In addition a much broader framework was introduced to capture many of the existing documents and formalise their input into the Asset Management Strategy. This was to be a fundamental aspect of the cultural change needed to move forward. A clear objective of the changes that was learnt from the previous structure was to re create the visible link between each employee's effort and the delivery of a common divisional goal. This transparency was designed to capture employee ownership and input and ensure it was seen as being valued. It was also seen as a way of restoring some of the employee satisfaction lost in the preceding years under the process structure where many skilled employees felt they had simply become production line workers rather than strategists or maintenance engineers.

By far the most significant change undertaken was to aggregate the key functions of the Asset Management Model into a single strategically focussed Asset Management division (Powercor Network Division) to ensure a consolidated view on the asset whole of life cycle prevailed when making decisions. The consolidation of a single management focus upon assets occurred. A lack of local ownership of the assets was becoming an emerging issue also at the time. A reliability audit undertaken by the Regulator also confirmed this was an emerging behavioural trend across the company. To address this issue and provide a level of local asset ownership six new Regional Asset Management roles were implemented. These roles also had a geographic responsibility to ensure all the Asset Management policies were being applied consistently and correctly. This also led to beneficial feedback being obtained and fed back into the policy reviews with a minimum of bureaucratic processes and paper work.

In any analysis of the value of maintenance programs the organisational structure needs to facilitate the continually linking of reliability performance to maintenance programs as an initial, and fundamental, step. To support this throughout the structure additional training in RCM II analysis and methodologies training has commenced for many Asset Management staff to strengthen the linkage between all activities undertaken and the delivery of reliability improvement outcomes. Currently the reactive nature of breakdown costs and preventative works could be substantially improved by being performed in a proactive and a planned manner. Breakdown works to restore electricity supplies presently incur an additional cost in the order of 120 %

to 175 % due to their unplanned nature and the inefficiency this introduces when compared to performing the task as a planned maintenance activity.

Significant reliability and employee satisfaction benefits were able to be delivered by addressing the cultural paradigm that was present and had evolved from separating all resulting output benefits of the maintenance program and maintenance engineers who were implementing the work schedule. The process structure had introduced an unpalatable scenario that needed resolution from a work ethic viewpoint. If employees feel their contribution towards maintaining the system is treated equally, regardless of the achievement of any assessable benefit or improvement, a large de-motivating and counter productive culture can, and in fact did, emerge. In this instance neither incentive nor recognition occurred for performing the maintenance program well. The return to a structure that linked tasks with outputs and measured benefits through actions that had been analysed and implemented as a part of a broad strategy delivered significant step improvement benefits as can be seen in figure 3.

Often engineering based organisations have difficulty in harnessing the full potential of their employees. Many believe their staff is their greatest asset yet the pursuit of employee satisfaction and consultation on items that would provide greater motivation for staff is seen as "softer" supporting functions rather than the core to delivering engineering excellence. Figure 3 below shows the impact of the revised structure and evolved Asset Management strategy upon reliability performance in 2001 through 2005. These results challenge the view that employee and structure related topics are soft and suggests that perhaps they should be viewed as more "core" engineering issues.

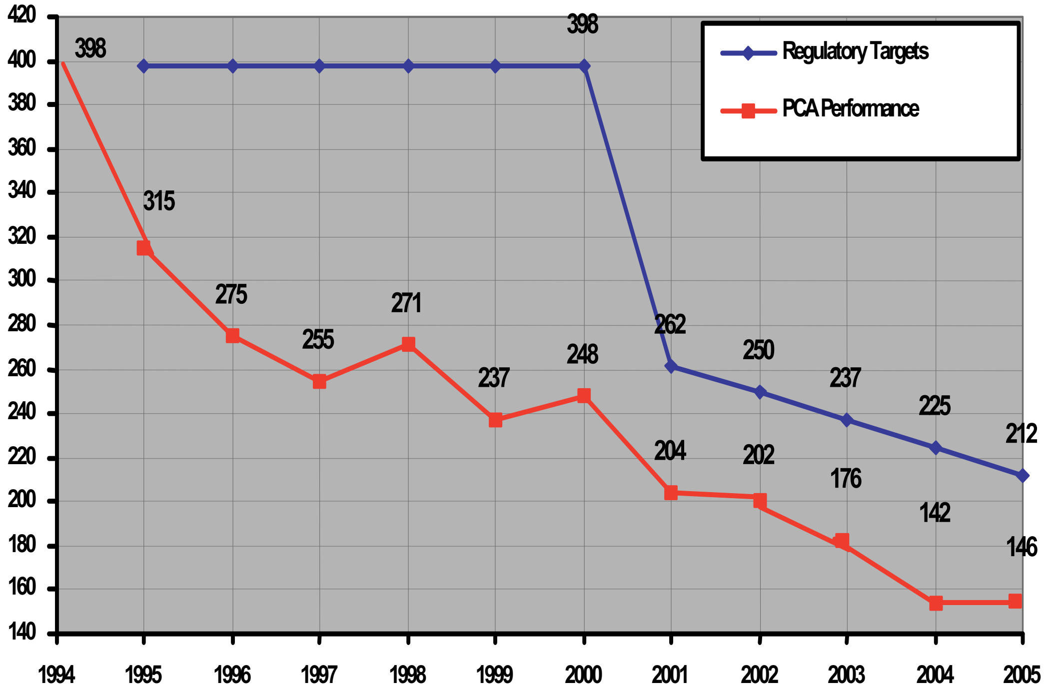


Figure 3: Powercor performance and targets

2001 - 2005 Step improvements and performance

Reliability performance 42% improvement on 2000.

Customer complaints down 50% from 2000

National leader in electrical industry in health and safety with a lost time injury frequency rate of 0.5 (per million man-hours worked).

A strong financial performance was delivered with better than of targeted performance in both Capital and Operation areas.

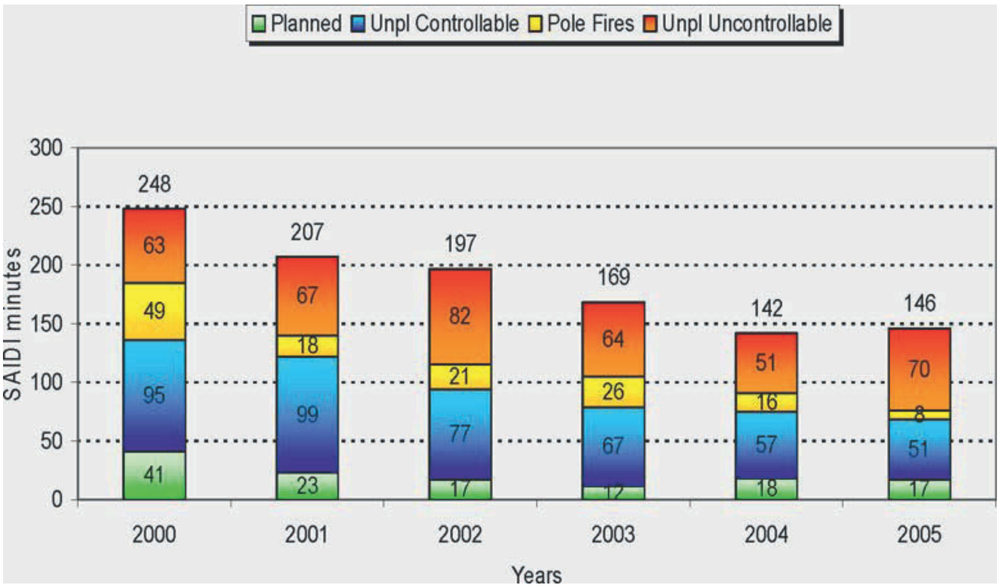


Figure 4: Powercor Reliability Performance 2000 vs 2005 by Cause and SAIDI Impact

2004 Consolidated Asset Management Delivery

Development of the Asset Management Strategy to date has largely focused upon preventing outages as a result of asset failures. In moving forward the efficient deployment of resources and completion of maintenance tasks needed to become a key consideration. This largely relates to the plan, schedule and execute the task activities as undertaken in the Maintenance / Construction Management module of the Asset Management model. (figure 2) Powercor's asset inspection and maintenance process is based around a rotating 5 year cycle.

The analysis of failures and considerations of the costing of undertaking work on a geographically dispersed asset has enabled Powercor to conclude that the efficiency gained through bring forward maintenance will be more than offset by the reduction in site visits and reduced break down attendances. Contrasting the present inspection process assets will continue to be inspected in a rotating 5 year cycle, however all works will be completed in year one with no further maintenance works required for the remainder of the cycle.

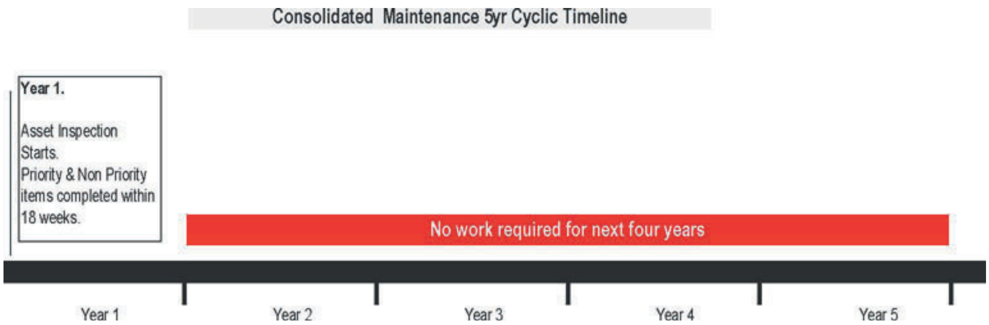


Figure 5: Powercor Proposed Consolidated Maintenance Practice Timeline

4.1 Improvement in the application of Diagnostic tools

A key component of the asset management model that it has an inbuilt allowance for flexibility and innovation in the improvement initiatives step of Continuous Improvement module (figure 2). Powercor has reviewed the reliability outputs delivered (figure 4) and recognised if an improvement in diagnostic techniques was possible then the sites of potential pole fires could be identified and preventative maintenance work undertaken prior to the failure occurring. If this additional work was undertaken simultaneously with normal consolidated maintenance process, as explained above, then an efficient solution to delivering the reliability improvement benefits would be achieved.

Pole fires can occur as a result of a pollutant building up on an insulator or cross arm surface. This build up may become conductive in the presence of moisture such as mist or light rain which in turn will provide a for leakage currents to flow from the energized conductors to earth. These currents, or tracking as it is commonly known, can flow on the surface and within a cross arm and will frequently concentrate on the un-earthed metal bolt that attaches the cross arm to the pole. This leads to the bolt becoming hot and a smouldering fire to commence leading ultimately to the cross arm failure and a power outage.

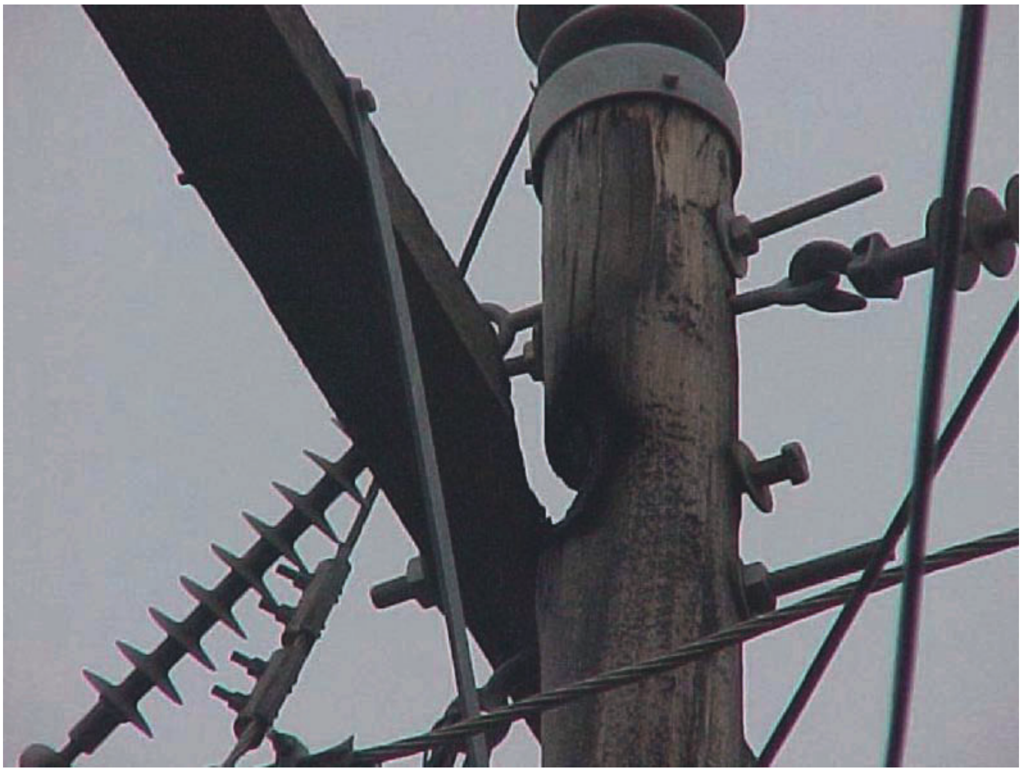


Figure 6: Pole fire

In 2004 Powercor evaluated the corona camera as a tool for detecting potential pole fire locations and other maintenance items. It was assessed that the camera had significant potential for detecting defective assets and in conjunction with thermographic inspection supports routine ground based asset inspection programs to mitigate the risk of earth leakage faults, television and radio interference and in particular pole top fires. A 3 year corona camera inspection contract was initiated. The technology is based on the detection of corona discharge via high resolution video camera "corona camera". The camera detects UV emissions from discharges resulting from ionisation of air caused by leakage current. This activity may also result in audio noise or radio interference which provides a greater breadth of scope for the diagnostic tool to perform additional functions as a part of customer initiated enquiries or complaints. The camera is able to detect severity of leakage current from the corona discharge, its location and potentially the likelihood of pole fire ignition or insulation failure. A sample report from the camera is shown below where the corona is clearly visible. (corona appears as white discharge around insulator)

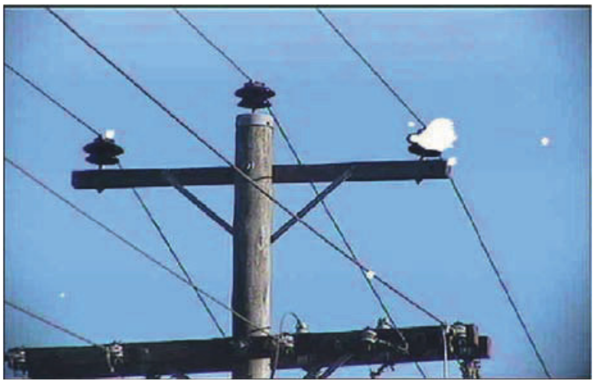


Figure 7: Sample Reports and Associated Analysis

Figure 8: Sample Reports and Associated Analysis

SAMPLE 1

REPORT No VF46
DATE 26-11-03

Nature of Fault

Wind Speed	Moderate	Humidity (%)	Low
Temp °C	20	Precipitation	None
General rating of conditions for camera effectiveness		Very Good	

EQUIPMENT POLE No21 LIS No4-22317 FDR CDN3

OBJECT 22KV FOG INSULATOR	Road side phase
DAMAGE Ionisation Cloud	20mm diameter
CAUSE	Cracked Insulator

ANALYSIS

Possible Cause

The level of corona discharge on this insulator compared with the companion insulators would indicate a damaged insulator rather than a pollution issue. If pollution was the cause, corona would be expected on all 3 insulators to varying degrees; however this can only be confirmed by a close inspection of the insulator. The damage to the insulator could be as obvious as a cracked or chipped shed or just deterioration in the insulator surface glazing which would only be detected by close inspection. In addition, the insulator has a lean which indicates a loose fitting on the cross arm which will increase the risk of pole fire.

Priority

Pole and Cross arm condition looks good therefore the risk of pole fire ignition would be low in this case, however as the cross arm deteriorates with time this is a potential pole fire risk location. In addition the insulator may generate TVI and under damp conditions the pole may become alive due to tracking.

Recommendation

It is recommended that the pole top structure be replaced as a priority 2 defect and the insulator retained for closer inspection.

As can be seen from the above sample report an accurate assessment can be made as to the likelihood of a pole fire occurring and an associated risk level can be applied to program the cross-arm for replacement. This pro active diagnostic technique has been implemented at Powercor and the benefits delivered in proactively managing pole fires can be seen in the reliability performance table shown in figure 2. Additional benefits have also been delivered through an improved ability to detect radio and television interference enquiries that are frequently raised by customers.

OPTIMISING FLOOR SPACE DURING RETROFITTING OF HIGH-RISE OFFICE BUILDINGS

Chintha Perera[a], Sujeeva Setunge[a], Arun Kumar[b] and Tom Molyneaux[a]

[a] *School of Civil and Chemical Engineering, RMIT University, GPO box 2476V, Melbourne, Australia.3001.*

[b] *CRC for Construction Innovation*

Abstract: A need for an efficient decision support tool for planning of office building refurbishment is emerging due to the significant increase in the number of ageing buildings in the world. In order to meet this need, all the aspects of building refurbishment projects such as waste minimisation and management, structural strengthening and floor space optimisation, estimation of residual life and construction management issues need to be integrated. Whilst there are number of such tools reported in the literature to cover other parameters in relation to re-life of buildings, they do not include evaluation of the optimising of rentable floor space and corresponding structural strengthening needs. This paper discusses the methodology for development of decision support frame work to compare and assess options available to the design team of a re-life project in optimising rentable floor space. The issues to be addressed during floor-space optimisation have been identified and prioritised through a Delphi process. These issues are then mapped against the life cycle of a refurbishment project and consolidated. Decision tool developed as a matrix to map existing as well as innovative structural strengthening techniques to strengthening needs is presented to demonstrate the application of the proposed model.

Key Words: Retrofit, Office buildings, Structural strengthening

1 INTRODUCTION

The significant growth in the construction of new commercial office buildings over the past 30 years has left a large stock of ageing buildings. The repair and maintenance expenditure of aged buildings is expected to increase due to the structural deficiencies and functional obsolescence. In such situations, the owner of a building has to decide the most judicious moment to undertake refurbishment to extend useful life of their buildings and improve their overall value or to replace them with a new building.

Over the past 5 years, an increasing number of projects are exploring the advantages of retrofitting of aged buildings and therefore, office building retrofitting market has seen considerable growth. The followings explanation of this paradox is given in Literature [1]:

- the life span of office buildings is much shorter than residential buildings;
- users' requirements have considerably changed in terms of office equipment, communications, automation, quality of use and comfort;
- buildings that do not offer all the amenities for comfort and flexibility are difficult to sell or rent;
- retrofitting a building costs much less than demolition and reconstruction;
- office buildings are classified amongst the buildings presenting the highest energy consumption.
- With the changes in legislation, approval of a new development of the same user capacity may not be permitted within the envelop of the old building

In developing a strong business case for re-life rather than demolition and re-build, a major issue to be addressed is maximizing of rentable floor space, which often puts a re-life project at a disadvantage. In increasing usable floor space, options available to the client include removal or adding floors and partitions, relocating services, cutting openings or extending floors and relocating lift wells etc. In each case, innovative strengthening schemes have much to offer when considering upgrading of the existing structure.

For a retrofitting project to be successful, the owner has to establish a suitable retrofitting plan of action. If a client makes inappropriate choices, the outcome may be a time and/or cost overrun and general dissatisfaction. Nowadays, a number of decision support tools are available to the owners to assess the current condition of their buildings with respect to deterioration, functional obsolescence, energy consumption and environment quality, before making a decision to retrofit their buildings. Those decision support tools also help to choose appropriate retrofitting actions and to estimate the cost. However, the available decision support tools do not cover the building structure. Consequently the "Floor space optimisation" strand of Re-life project funded by CRC for construction innovation is looking into the development of a management decision support tool that assists the owners in selecting an appropriate strengthening scheme during optimizing of floor space in a re-life project.

2 OPTIMIZING OF FLOOR SPACE DURING RETROFITTING OF BUILDINGSS

It is a significant challenge for engineers to satisfy the clients' requirements during retrofitting; specially optimizing the usable floor space. Usable floor space of existing building can be optimized by modifying the layout of an existing building such as removing or adding of floors and partitions, relocating services, cutting openings or extending floors and relocating lift wells etc. In these situations, innovative structural strengthening schemes could be implemented to strengthen the existing structure. Strengthening of an existing structure may also become necessary under the following circumstances:

1. When buildings are exposed to harsh environments such as de-icing salts, chemicals or air-borne salt spray, or have inappropriate details, the structures may experience significant deterioration in the form of steel corrosion, concrete cracks and spalls [3]. Such deterioration may result in structural inadequacies that adversely affect the structure or its members,
2. In situations where the building owners make decision to change the usage of a type of building depending on the business demand in the current market environment. As a result, the existing structure of a building may or may not be strong enough to withstand the new loading,
3. Structures constructed in early days may have been designed to carry loads that are significantly smaller than the current needs, possibly due to increase demand usage [3],
4. . Structural inadequacies may arise due to errors in initial design or construction and changes of design standards.

By carefully implementing appropriate structural strengthening methods, re-life of buildings can be justified/achieved rather than adopting demolishment and re-building

3 STRUCTURAL IMPLICATIONS OF OPTIMISING USABLE FLOOR SPACE

In situations where the use of the office floor requires changes to services such as lift wells air ducts, ventilation, lighting and the type of loads applied on it, load transfer mechanism of the building structure may require strengthening or re designing and re-building.

Selection of an appropriate strengthening method is dependent on the materials of construction. Structural elements of existing buildings are commonly constructed of concrete, steel, timber and masonry. Concrete is the most commonly used building material and widely used in the forms of in-situ concrete, precast concrete and post-tensioned concrete.

The structural strengthening can be achieved by section enlargement, external post tensioning, external bonded steel elements, bonded advanced fiber reinforced polymer (FRP) composites, span shortening, or a combination of these techniques [2,3,7,9,12-26]. No matter what strengthening technique is used, the ability to perform as an integrated system can be obtained only by providing an adequate bond between the existing member and the repair/reinforcement to ensure monolithic structural behaviour [2]. Stress concentrations resulting from added material should be investigated as they may cause a localized failure.

Unfortunately, there is no specification or design that covers all repair/upgrade scenarios and engineers, architects, and contractors must be innovative and thorough, in their design details, specifications, and applications.

4 MANAGEMENT SUPPORT TOOLS AVAILABLE FOR RE-LIFE PROJECTS

A good retrofitting action plan will lead to the success of a re-life project. With the aid of decision-making tools, it is possible to select the most suitable retrofitting action [3, 4, 5, 27, and 28]. A number of such tools have been developed for office buildings with TOBUS, MEDIC, EPIQR and INVESTIMMO being the most commonly reported in the literature [3, 4, and 5].

The decision-making software tools, EPIQR (for apartment buildings) and TOBUS (for office buildings) have been developed for the assessment of retrofitting needs of buildings in European countries. The use of these tools can facilitate a quick and accurate diagnosis of the condition of the existing building in terms of its major area including construction, energy performance, indoor environmental quality, and functional obsolescence. The main advantages of using these tools are the ability to evaluate various refurbishments and retrofit scenarios, and cost of induced works, in the preliminary stages of a project [1, 6].

In EPIQR and TOBUS, deterioration of building materials and components are described by the use of a classification system with four classes. The prediction of the period of passing into the next deterioration state is of high interest as this is directly connected to higher refurbishment costs. The prediction of qualitative deterioration states are important and correspond to key moments in the element's life where some refurbishment action has to be taken [1,6].

European countries have used another software tool entitled MEDIC to predict the future degradation state of building. MEDIC is intended for use with EPIQR and is based on a subdividing of the building into 50 elements. MEDIC calculates the remaining life span of a building element not as a deterministic unique value but as a probability distribution. It can help the owner of a building to decide the most judicious moment to undertake refurbishment to achieve his short and long term financial needs [3, 8, and 29].

Following the footsteps of EPIQR an TOBUS, a decision-making tool for long term efficient investment strategies in housing maintenance and refurbishment – INVESTIMMO has been developed in European countries. It has been aimed at evaluating housing maintenance and refurbishment options, which covers expectations of tenants, housing market, and quality of building upgrading and environmental impacts in addition to the factors identified in TOBUS [6].

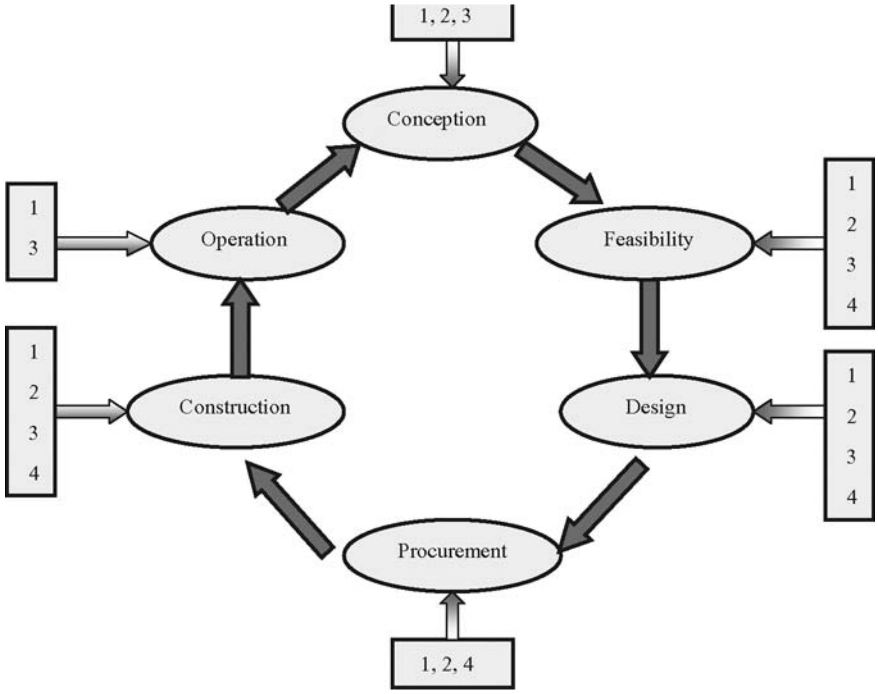
From the review of literature, it was clear that these tools can be used to evaluate the general state of buildings with respect to some of the aspects of building re-life projects such as service life, functional obsolescence, energy consumption and environment impacts. However, no reported work presents a decision support tool, which can be used to compare and assess options available to the design team of a re-life project in optimizing rentable floor space.

5 INTEGRATED FRAMEWORK FOR DECISION MAKING

In order to develop a decision support tool for selecting the most appropriate structural strengthening scheme during optimizing of floor space, the preliminary objective was to identify the clients' requirements and problems associated with structural strengthening. Through the review of literature and based on the experience, a list of issues, which may influence strengthening work has been compiled. Consequently, Delphi process was utilised to rank them based on the importance in refurbishment projects [10, 11]. The following issues were identified as important by the experts engaged in research/practice in the field of building refurbishment:

1. change of use of floors,
2. cutting openings in floors and extending floors,
3. relocate/renew services,
4. structural appraisal prior to refurbishment and
5. safety reliability issues in structural strengthening.

The next challenge faced by the research team was integrating the issues specific to structural strengthening with other issues such as construction management, estimation of the residual life and managing waste and recycling. This was achieved by mapping the issues against the project life cycle as shown in Figure 1, established through an industry workshop with the partners of the project. This identified the input parameters required at each stage of the project life cycle. In this paper we are focussing on the floor space optimisation and structural strengthening during building refurbishment. Detailed mapping of issues to different phases of the project are captured in Table 1



Where 1, 2, 3 and 4 are the inputs from main aspects of building refurbishment projects

1 - Construction management issues,

2 - Structural strengthening and floor space optimisation,

3 - Estimation of residual life,

4 - waste minimisation and management.

Figure 1: Life cycle of a building refurbishment project

Table 1: Mapping of floor space optimisation issues to different stages of the project life cycle

Phases of life cycle of a building / Issues Associated with optimising floor space and structural strengthening	Conception	Feasibility	Design	Procurement	Construction	Operation
Structural Appraisal	√	√	√			
Change of use of floors	√	√	√			
Relocate/renew services	√	√	√	√	√	
Impact on structural strength		√	√			
Safety issues for structural strength		√	√			
Cutting opening/extending floors			√		√	

6 INPUT PARAMETERS TO THE LIFE CYCLE OF BUILDING REFURBISHMENT

Expanding the issues presented in table 1, researchers identified detailed decision parameters required at each phase of the project life cycle to achieve an optimised solution for usable floor space. These were identified through extensive industry consultation complemented by a review of published work. Decision parameters are given in Table 2. For example, at the project conception stage most important parameters are remaining life of the building, structural capacity of the existing building, and structural constraints in the existing building such as flexibility in altering the structure, changes to loads or load transfer mechanisms and structural requirement in relocating services. These issues are identified by studying in-detail the refurbishment reports of two case-study buildings: Council House 1 of Melbourne City Council (CH1) at little Collin Street in Melbourne and Health building at George Street in Brisbane. The two case study buildings covered two distinct refurbishment scenarios. CH1 building refurbishment was driven by a need to arrive at 6-star energy rating to match that of the adjacent council building CH2 which is a state-of-the-art green building. This required drilling through the flat slab floor to relocate services, opening up of light wells to improve natural lighting and ventilation and changes to the windows. George St. building in Brisbane has utilised a concept of integration of two adjacent buildings to optimise the floor space. Through the integration, floor space has been optimised by removing two sets of service cores and demolishing a stair well and using a common lift lobby. For both case study buildings the evaluation methodology for the parameters given in table 2 has been compiled by the research team. This information can be presented as a decision support tool, which will allow the project team to establish typical values for each of the parameters and ascertain the feasibility of a given floor space optimisation option. A typical example is given in table 3 and explained in section 7.

Table 2: Parameters to be considered in developing guidelines for optimising floor space and structural strengthening of a building

Phases of life cycle of a building	Parameters to be considered developing guide lines for optimising floor space and structural strengthening of a building
Conception	● remaining life of the existing building ● structural capacity of the existing building ● structural constraints ● change in load conditions ● change in load transfer mechanisms ● possible options for floor framing and relocation of services ● services to be considered for relocating or renewing ● structural requirements to support relocation of services
Feasibility	● cost of different structural treatment options, floor framing and relocating/renewing services ● availability of technology and expertise for structural appraisal, change of use of floors and relocate/renew services ● reliability of structural appraisal ● environmental issues on structural appraisal and change of use of floors ● user comfort due to relocate/renew services ● consider innovative options for relocating/renewing services ● integrate or evaluate structural options with other issues ● structural safety margins
Design	● calculation of residual structural capacity of existing building ● specialist evaluation techniques ● calculation of structural capacity with new loading condition ● consider innovative options for relocating/renewing services and cutting openings/extending floors ● consider all possible safety requirements ● consider floor framing option which has a relatively low effect on overall structural strength
Procurement	● availability of expertise for relocating/renewing services ● Availability of expertise for structural strengthening ● contractual terms for relocating/renewing services ● reliability of relocating/renewing services
Construction	● proper planning of relocating/renewing services ● techniques adopted for cutting opening/extending floors ● Structural strengthening
Operation	–

Table 3: Matrix of mapping solutions to problems (Contd.)

Structural Strengthening Solutions	Application Projects	Characteristic	Possible Applications	Issues Associated with the solutions	Companies/Contractors
1. **Externally bonded composite system** **- Fiber Reinforced Polymer (FRP)** **- Hardwire® Steel Reinforced Polymer (SRP)**	• Utility tunnel at a University in South Florida • Restoration of swimming pool roof structure, Kalmthout, Belgium	• speed and ease of installation • lower cost • aesthetic appeal • light weight • non-corrosive • excellent fatigue behaviour • non-conductive • can be used for flexural, shear, and axial (confinement) upgrading and crack control • due to thin profile, r/f can be easily run in two directions for two way slabs • allow for carpet, tile, and other flooring finishes to be installed over the system without any significant change in floor elevation	• concrete slab, beam, column and wall • wooden beam, column • masonry	• can damage the externally bonded composite • problematic in flooring applications when used on cracked concrete slabs as these cracks may allow reflective cracking throu' polymer topping.	• VSL International Ltd • Structural Preservation System (USA) • Edge Structural Composites (USA) • C.A. Lindam Companies (USA) • Watson Bowman Acme Corp. (USA)
2. **Section Enlargement**		• relatively easy method • not good in harsh environment. Possible corrosion of embedded reinforcing steel. • can increase load carrying capacity or stiffness • additional concrete can be placed in the form of an overlay or jacket	• concrete columns, beams, slabs	• Increase the weight of existing structure. weight can be minimised	• VSL International Ltd • Structural Preservation System (USA)
3. **External Post-tensioning**	• Two-span steel truss bridge (48-48m) over River Aare at Aarwangen, Switzerland • Pier 39, Parking Structure San Francisco, USA • Double-tee stems on an overpass located on the premises of a University in Washington. Damaged due to overheight truck.	• Possibility of controlling and adjusting the tendon forces, inspecting corrosion protection and replacing tendons. • can be used for both reinforced and prestressed concrete • minimal additional weight to the repair system • ecconomical • requied less time to complete • can be used for flexural , shear and axial (confinement) upgrading	• Applicable for structural steel, composite steel-concrete, timber and masonry structures. e.g. building with masonry walls, girders in buildings, roof structures, circular structures such as silos, reservoirs and large masonry chimneys. • great success to correct excessive deflections and cracking in beams and slabs, parking structure and cantilever members. • due to minimal additional weight of repair system, this is effective and ecconomical for long span beams • existing concrete structures against fatigue and cracking	• vulnerable to corrosion, fire and vandalism. However, improved ductility & fire proofing can be achieved by placing reinforcement in ducts that can grouted after stressing of tendon. Protection can also be achieved by encasing the post tensioning system in concrete or by using shotcrete. • Externally bonded bars can be damaged by traffic. However, that can be prevented by installing the system in grooves made in existing member. • requies access to sides and sometimes ends of member	• VSL International Ltd • Structural Preservation System (USA) • C.A. Lindam Companies (USA)
4. **Bonded steel element**	• Quatier des Celestines, a 19th century building in Namur	• Steel element can be steel plates, channels, angles or built-up members • steel elements can be bonded with epoxy adhesive. In addition to epoxy, mechanical anchors can be used to ensure steel element will share loads in case of adhesive failure. • can improve shear and flexural strength	• Steel plate bonded to tension face of concrete beam can increase flexural capacity, flexural stiffness and in deflection and cracking decreases • Steel plate bonded to side of the member can improve shear strength of concrete beam. • install structural steel beams on the underside of structures and steel brackets at column heads. • long-term solution for structures subjected to aggressive environment	• influence to corrosion and fire.Suitable corrosion protection system and fire protection system can be used. • due to their restrictive length, steel elements need to be spliced which complicate the design and construction operation • existing reinforcement can be damaged while placing the anchors. Considerable site work is required to accurately locate existing reinforcement. • Expensive false work is required to maintain the steel work's possition during bonding. • Adding strengthening material under existing structure may redue the usable headroom	• VSL International Ltd • Structural Preservation System (USA) • C.A. Lindam Companies (USA)

Table 3: Matrix of mapping solutions to problems

Structural Strengthening Solutions	Application Projects	Characteristic	Possible Applications	Issues Associated with the solutions	Companies/Contractors
5. **Span Shortening**	• Shortened span in Parking Garage	• reduce the force in overstressed beam • increase the load carrying capacity • best material for this application is steel, which is quick to install.		• result in loss of space and reduced headroom. • New footing for new column is expensive less expensive approach is to install diagonal braces that extend from the bases of existing columns	
6. **Hybrid strengthening method** eg. steel/CFRP	• Strengthening of a roof system of an elementary school in New Jersey. They wanted to install skylight on existing roof slab. • Transformation of former school building to library • Roof structure of warehouse at Brussels in Belgium (due to calculation error,capacity of main supporting beams are not suticient)	• less expensive • aesthetically pleasing • fast application	• slab with opening • rib slabs		• Structural Preservation System (USA)
7. **Removing existing concrete and casting with lightweight concrete overlay (with steel wire mesh or steel bar)**	• Baltimore's historic Hippodrome Theater renovation -France	• increase bending capacity by increasing effective depth of existing bottom r/f • at support, embedded steel steel r/f in overlay increase the bending capacity • r/f in overlay limits cracking of overlay	• concrete slab and beam	• adequate surface preparation is required to ensure the bond between overlay and existing structure • overlay replace existing topping and therefore small increase in dead load, which offset by using lightweight concrete.	• Structural Preservation System (USA)
8. **Hardwire**	• Baltimore's historic Hippodrome Theater renovation -France	• hardwire made of ultrahigh strength steel wires twisted together to form reinforcing steel cords wires bonded to existing structures with epoxy adhesive • hardwire works as additional r/f to provide tensile strength	• reinforced and prestressed beams, girders, and slabs to provide additional flexural strength. • can be used on sides of beams and girders to provide additional shear strength. • can be used on structural elements to increase their capacity.		• Structural Preservation System (USA)
9. **Hardwire steel belt with polymer flooring system**	• Leading national home improvement warehouse • multi-story warehoue/ distribution facility for a large automative parts distributor in illionis	• hardwire does not replace today's polymer flooring system. However, additon of hardwire dramatically strengthens retrofitflooring design capabilities. • combination of polymer flooring with these thin steel belts creates a system with the ability to increase structural capacity, blast resistance and general toughness of concrete • ability to integrate with multiple polymer and cementitious materials	• heavy trafficked warehouses, manufacturing slabs, elevated slabs, slabs on pile construction, and slabs that are structurally deficient from improper design or construction • viable solution where emission and odors are an issue.		• Structural Preservation System (USA)
10. **Replacing existing wire system with encapsulated monostrand post tensioning system**	• Parking Garage in New York	• existing structure to be post tensionned system • time and cost effective			• Structural Preservation System (USA)
11. **Epoxy mortar**	• Timber floor in a castle	• can be used to strengthen old timber structures. excellent material to repair wood because it has strong bond to wood and MOE is nearly equal to wood.	• Timber - A large groove was sawed on top side of the beam. Steel r/f has been placed in these groove. Then grooves are filled wth epoxy mortar.		

7 EXAMPLE OF A DECISION SUPPORT MATRIX FOR STRUCTURAL STRENGTHENIGN ISSUES

Table 3 demonstrates a typical matrix developed to provide decision support for structural strengthening issues at the conception and feasibility stages of a re-life project. The matrix covers possible structural strengthening solutions, typical application projects, characteristics of the strengthening method, possible applications and issues and the local expertise availability for a typical scheme. This allows the project team of a re-life project to identify possible strengthening solutions and feasibility of these prior to engaging a consultant for a specific issue. The complete tool covers other information such as typical costs and sample calculations for a typical strengthening solution. This matrix has been calibrated using data from the two case study buildings mentioned earlier: CH1 building in Melbourne and 63 George St. Building in Brisbane [30]. Similar matrices are currently being developed to cover parameters identified in table 2 such as options in improving energy rating using relocation/renew services, façade and window treatment and improved ventilation, integration of structural issues with other issues such as waste minimisation and construction and procurement methods.

8 CONCLUSIONS

The paper presented the challenges faced by engineers in decision making during re-life of buildings. It then presented a methodology for identifying major issues affecting optimisation of floor space and structural strengthening of buildings during a refurbishment project. A method for mapping major issues identified as influencing re life of buildings to different stages of the project life cycle has been presented. This model presents decision makers with valuable guidance at different stages of the project life cycle of a re-life initiative. An example of a detailed decision support guideline developed to cover one major parameter has been presented to demonstrate the application of the model. This guideline given as a matrix in table 3 has been validated with data from two case study buildings and other published work.

9 REFERENCES

[1] Caccavelli, D. and Gugerli, H., 2002. "TOBUS-a European diagnosis and decision-making tool for office building upgrading", Energy and Buildings, 34, pp:113-119.

[2] Rey, E., 2004. "Office Building Retrofitting Strategies: Multi-criteria Approach of an Architectural and Technical Issue", Energy and Buildings, 36, pp: 367-372.

[3] Flourentzou, F. et al., 2000. " MEDIC - a method for predicting residual service life and refurbishment investment budgets", Energy and Buildings, 31, pp:167-170.

[4] Allehaux, D. land Tessier, P., 2002. "Evaluation of the functional obsolescence of building services in European office buildings", Energy and Buildings, 34, pp: 127-133.

[5] Brantz, E. and Rasmussen, M. H., "Assessment of building conditions", Energy and Buildings, 34 (2002) pp 121-125.

[6] Balaras, C.A., 2002. " TOBUS - A European method and software for office building refurbishment ", Energy and Buildings, 34, pp:111-112.

[7] Steelcase, Workplace Issues: One in a Series. " Reclaiming Buildings – Strategies", available online http://www.steelcase.com/Files/de9c9b911ac84b3bbdc25f352373316a.pdf (downloaded: February 2005)

[8] Flourentzou, F. et al., 2000. " MEDIC - a method for predicting residual service life and refurbishment investment budgets", Energy and Buildings, 31, pp:167-170.

[9] News from the Cement and Concrete Industries, Concrete Monthly. "Design, application techniques key to successful structural repair, strengthening of aging concrete facilities", Available online http://www.concretemonthly.com/monthly/art.php/516 , February 2004 Issue.

[10] Chan, A. P. et al. 2001. "Application of Delphi method in selection of procurement systems for construction projects", Construction Management and Economics, 19, pp:699-718.

[11] Gordon, T.J., 1994. "The Delphi Method", AC/UNU Millenniums Project.

[12] Alkhrdaji, T. & Thomas, J., "Upgrading Parking Structures: Techniques and Design Considerations", available online, http://www.structural.net/News/Media_coverage/media_csi.html , (downloaded: February 2005)

[13] VSL Strengthening Systems Brochure, available online http://www.vsl.net/downloads/VSL_Construction_Systems/Sections/Strengthening.pdf, (downloaded: February 2005)

[14] VSL Report series 1, "External Post-tensioning-Design Considerations, VSL External Tendons, Examples from Practice", available online http://www.vsl.net/downloads/VSL_technical_reports_PT_External.pdf , (downloaded: February 2005)

[15] Watson Bowman Acme Brochure, "Composite Strengthening Systems", available online http://www.wbacorp.com, (downloaded: February 2005)

[16] D. S. Brown Brochure, "Fiberbond FRP Structural Strengthening System", available online http://www.dsbrown.com/downloads/pdfs/Brochures/FRP/FRPbrochure.pdf, (downloaded: February 2005)

[17] C.A. Lindman Brochure : Exterior Restoration and Structural repair, available online http://www.calindman.com/htmlfiles/services1.htm1, (downloaded: February 2005)

[18] Edge Structural Composites Brochure: Strength through technology, available online http://www.edgest.com, (downloaded: February 2005).

[19] Brosens, K. et al., 2001, "Strengthening of concrete structures with externally bonded reinforcement - Design philosophy and practical applications", 10th International Congress on Polymers in Concrete, Honolulu, Hawaii, USA.

[20] Van Gemert et al., 2004, "Strengthening of concrete constructions with externally bonded reinforcement: design concept and case studies", IMTCR 2004 innovative materials and technologies for construction and restoration, University of lecce, Italy.

[21] Morgan, D. R., 1996. "Compatibility of concrete repair materials and systems", Construction and building materials, 10, 1:pp57-67.

[22] Setunge, S., Nezamian, A. & Lokuge, W., 2004. "User friendly guide for rehabilitation or strengthening of bridge structure using Fibre reinforced Polymer Composites", Report 2002-005-C-04.

[23] Grace, N. F., 2004. "Concrete Repair with CFRP- Evaluating the durability of externally bonded fiber-reinforce polymer plates and fabric exposed to the environment", Concrete International, pp:45-52.

[24] Einde et al., 2003. "Use of composites in civil structural applications", Construction and Building Materials, 17, pp:389-403.

[25] Toutanji, H. A. & Gomez, W., 1997. "Durability Characteristics of concrete beams externally bonded with FRP composite sheets", Cement and Concrete Composites, 19, pp: 351-358.

[26] Karbhari, V. M. & Zhao, L., 2000. "Use of composites for 21st century civil infrastructure", Computer methods in applied mechanics and engineering, pp: 433-454.

[27] Dascalaki, E., & Balaras, C. A., 2004. "XENIOS-a methodology for assessing refurbishment scenarios and potential of application of RES and RUE in hotels", Energy and Buildings, 36, pp: 1091-1105.

[28] Lam, K. P. et al, 2004. "SEMPER- II: an internet-based multi-domain building performance simulation environment for early design support, Automation in Construction, pp: 651-663.

[29] Linstone, H. A. & Turoff, M., 1975. "The Delphi Method Techniques & Applications", Reading, Mass: Addison-Wesley Pub. Co., Advanced Book Program.

[30] Perera, C. et al., 2006. "Structural strengthening for optimizing floor space during retrofitting of high-rise office buildings", Proceedings of the Clients Driving Innovation: Moving Ideas into Practice- Cooperative Research Centre (CRC) for Construction Innovation (12-14 March 2006)

CRITICALITY IN ASSET MANAGEMENT

Healy Dr. J.A.

JAHCon Physical Asset Management Pty. Ltd.

Abstract: This paper discusses the development and application of a systematic and auditable risk based Criticality Ranking for use in complex plant (both equipment and systems) and is based on work carried out for several organizations in Australia.

Criticality is often mentioned in texts and conferences dealing with Asset Management and Maintenance however little discussion takes place about its uses and applications and the important role it plays in facilitating many Asset Management functions. Limited resources and increasing consequences of failure in modern plant make objective and auditable Criticality assessment of known and likely failures increasingly important.

The Criticality assessment is a fundamental component of effective Asset Management and if not formally developed and implemented, an informal ranking will be applied by individuals even if it is not recognised as such.

A well developed Criticality ranking is beneficial to the organisation if it impacts positively on decision making and provides an objective guide when establishing appropriate responses for various end-user scenarios.

This paper outlines the development and implementation of a Risk based assessment of Criticality and discusses the application of Criticality in day-to-day Asset Management and Maintenance activities.

KEYWORDS: Criticality, Priority, Asset Management, Maintenance, Effectiveness, Efficiency, Reliability

1 1. INTRODUCTION

In the words of Russell Baker "The role of inanimate objects is to resist man and ultimately defeat him". Unacceptable outcomes from system or equipment failures usually arise from a lack of appreciation of the risk associated with operating these systems or equipment. The likelihood and consequences of a potential failure are considered in many instances only after a significant failure (or near miss) triggers a retrospective assessment of a risk the organisation had been living with for some time, often in complete ignorance. Rarely do these failures occur as a result of malicious damage therefore blaming employees or staff for failing to consider these possible outcomes is more of an indicator of poor management than poor employee performance. The only practical approach available for a modern organisation is to carry out a planned and structured assessment of the known and likely risks associated with the operation of the plant and instigate risk management strategies to limit the negative impact of these events.

Criticality based on risk assessments have application in many areas of the organisation and while this paper discuss the application of Criticality Assessments to Maintenance and more broadly to Asset Management, it is important not to overlook the need for similar studies in other areas of the organisation.

Starting from the premise that there is no way to completely eliminate all risk from the organisation, we must address the potential for negative impacts in a systematic and scientific way rather than wait for an event to occur to draw our attention to the risk. While the range of risks in a modern plant is enormous, fortunately not all are of equal significance and usually only a relatively small percentage requires the highest level of risk management to control them. In practice, only about 10% of the assets will have 'High' criticalities with the bulk of the assets having 'Medium' and 'Low' level criticalities. This does not mean that we need only assess the 'High' criticality items as we have no objective way of identifying which are the 'High' Criticality assets until we have assessed them all. This means that the Criticality assessment has to be carried out for all assets and systems in the area of interest before the 'Critical Few' can be identified from the rest.

Without a structured and systematic approach there is little likelihood of addressing all major risks in a useful time. Some of the potential risks will have limited impact on the organisation should they occur and therefore can be managed in a way that accepts some failure occurrences. Other failures can lead to injury or death or complete failure of the business and therefore have to be managed in a way that prevents all occurrences. The varying levels of risk identified by the assessment will require a range of responses commensurate with the risk level identified. Clearly a 'One-size-fits-all' approach is unlikely to be either cost effective, or effective in managing the risks identified.

2 CRITICALITY ASSESSSMENT

Criticality assessment is a structured methodology that provides a proactive approach for the assessment of risks in the organisation. Different approaches are possible while all have the same desired outcome – namely a relative or absolute ranking of the systems and/or assets by their 'importance' to the organisation. A comprehensive Criticality assessment would provide a ranked list of systems and/or equipment (a Criticality Ranking) identifying possible failure mechanisms, underlying failure rates, likelihood of impact from the failure and the consequences of the failure. The outcome of the assessment would typically lead to risk management strategies for equipment which would benefit most from the application of a particular management, operating or maintenance strategy. Criticality ranking also assists in making decisions about which equipment is most important for planning and scheduling maintenance work as well as document control and other management functions.

Criticality assessments can be carried out using various techniques varying from the 'Wild Guess' to the fully 'Quantitative Risk Analysis'. The more common systematic approaches to identifying and representing Criticality include 'Qualitative Assessments', 'Semi-Quantitative assessments' and 'Quantitative Assessments'. Within each of these broad groupings a wide range of approaches are often used. Each approach has its merits and disadvantages and should be clearly understood by those undertaking the study before committing to a specific approach. For an initial assessment of criticality, or where information is limited upon which to base the assessment, a Qualitative or Semi-Quantitative approaches is likely to be the most practical and can give effective results if well implemented. It is likely that the initial assessment would highlight the lack of suitable data as being a risk factor and the management of that risk may involve improved data collection and analysis in the future.

Quantitative assessments rely heavily upon historical data, or well based experience, to 'calculate' the rankings assigned to the equipment or failure mechanisms being assessed. These calculations typically include Probability of failure and Cost of each failure event. Obtaining the necessary data can be very difficult and often the data has to be pre-processed to get it into a form suitable for the analysis and any pre-processing must be understood and carefully carried out to prevent distortions of the final outcomes. It is obvious that where the failure mechanisms do not lend themselves to easy quantification that the Quantitative approach can be unsuitable. This is especially so when assessing the Consequences for Safety and Environmental failure mechanisms and despite claims to the contrary, quantifying a human life or serious injury or for that matter serious environmental damage, is unacceptable to the wider community and is impossible to defend to reasonable people.

It is important to select an assessment approach in keeping with the needs of the organisation as use of a too basic approach can result in insufficient resolution or rigor to implement the necessary risk management responses while an overly detailed approach can increase the cost and time for the assessment and result in a lot of redundant or unnecessary data being generated.

The following steps outline the practical audit process.

1. Consider each asset in its operating context – what it does & where is it?
2. Develop an equipment failure description – i.e. how can it fail?
3. Estimate the consequences of each failure – i.e. what happens when it fails?
4. Determine likelihood of each failure – i.e. how often is it likely to happen?
5. Calculate the Risk from the Likelihood & Consequences.
6. Sort the risk scores in decreasing order of importance.
7. Map the maximum risk score for the unit onto the 'A', 'B', 'C' scale (optional).

Individual systems or equipment can have more than one Criticality Ranking depending on the number of criteria being considered in the assessment. The choice of criterion is crucial as omission of important criterion can result in some significant failure mechanisms being left un-assessed [1]. While there are many criteria that can be included in a Criticality assessment there is also a need to minimise the criteria considered to keep the assessment workload within reasonable bounds. In practice, some common Criticality Ranking criteria include:

- Safety
- Environment
- Asset Damage (secondary damage)
- Production Loss
- Company image

By considering these (and possibly other) criteria and assigning a Criticality Score to the various failure mechanisms for each criterion it is possible to build up a Criticality Profile for the system or machine being assessed. In practical applications it is common to use the highest Criticality Score for a system or machine to represent the Criticality Ranking of the system or machine as a whole. The overall 'Criticality' of the system or machine can be used to support decision making throughout the organisation including projects, maintenance, operations, management etc.

In practice, the scale/s used for the assessment are more detailed than the final Criticality Ranking scale and therefore the 'Criticality Scores' can be 'mapped' onto a lower resolution 'Criticality Ranking' scale (typically 'A', 'B', 'C' or similar) to make it easier to use the Criticality Ranking on a day-to-day basis. The 'mapped' scores are then inserted into the CMMS to allow Criticality to be assigned to jobs for planning and other equipment management purposes.

Depending on the level of detail included in the Criticality Analysis, the results allow many Asset Management decisions to be made on objective criteria and also provide an audit trail for future review and updating. Some of the applications of the criticality analysis results include the following.

- *Provide a common communication tool.*
- *Process system design.*
- *Redundancy – equipment and systems.*
- *Maintenance programs.*
- *Work management based on known risk.*
- Identify spare parts requirements.
- Target maintenance budgets.
- Identify training requirements.
- Identify skill based risk.
- Quantify the maintenance commitment KSF's, KPI's.
- Provide a basis for system modelling.

3 CRITICALITY ASSESSMENT PROCESS

A Criticality Analysis can be carried out by applying Risk Analysis techniques to identify the Risk associated with particular equipment failure mechanisms in a specific operating context. AS/NZS 4360:1999 is an established standard for the calculation of risk and can be useful in the assessment of risk based criticality. The risk can be calculated by applying a score to the likelihood of an event occurring and another score to the expected consequences associated with each occurrence. These scores are multiplied together to give a risk score which can be used as a measure of criticality. The risk (or criticality) ranking is used to identify the most important assets in the organisation. Depending upon the range of scores available (1...5 each for Likelihood and Consequences in this example) and whether the scale is ascending or descending with increasing importance, the Risk ranking can range from a score of 1 through to 25. In the examples in this document the scales range from 1...5 and increasing scores correspond to increasing importance. Therefore the Risk ranking can range from a low score of 1 through to a high score of 25.

4 DETERMINATION OF CRITICALITY (C)

Each asset has one or more primary functions in a process and can have one or more secondary functions, some of which may require more sophisticated operational or maintenance activities than the primary functions. In some circumstances, the criticality of the secondary functions can be more important than the primary function. Identifying the critical failure mechanisms (strictly speaking failure modes) and developing effective strategies to manage these is the role of the criticality assessment. Table 1 shows a simple assessment sheet used to identify the Criticality of some equipment.

An objective analysis will identify the critical items and ensure resources are not wasted on items that add little value to the process. These include low cost items that have built in redundancy or have little or no effect on production, safety etc. This allows the maintenance team to concentrate on the most important functions (primary & secondary) performed by the asset. It is very important when establishing the assessment project that adequate consideration is given to the practical aspects of the project especially how to ensure uniformity of response and consistency of approach. In this instance uniformity is taken to mean a common result from similar inputs by different teams at a point in time while consistency relates to getting the same response over time regardless of fatigue or other factors.

Very often the assessment methodology will change over time as experience is gained, fatigue sets in and resources become difficult to obtain. Any legitimate change to the methodology should be closely controlled using authorities and approvals and formally documented. Any equipment assessed under the earlier methodology must be reassessed to reflect a significant change in the methodology. Without this ability to review equipment previously assessed the overall criticality assessment would be undermined by the use of different assessment methodologies for different equipment.

The best assessments are made by those with appropriate experience, a systematic approach and using a well designed methodology. In addition, uniformity and consistency between teams can be achieved by use of standardised and agreed look-up tables for Likelihood and Consequences and by use of trained facilitators who can guide the teams in how to interpret the assessment methodology as well as share learning's from other teams. The facilitator also helps to 'drive' the assessment programs selecting assets for assessment, identifying suitable team participants, setting up meetings, chasing up necessary information such as manuals and drawings and ensuring that any follow up work is carried out. Review stages after the initial assessment are often used to authorise changes to the criticality ranking of equipment, provide support to the teams and ensure a high level of assessment is achieved at all times.

Table 1

Example of a very basic Criticality assessment sheet.

				PRODUCTION LOSS				SAFETY				ENVIRONMENT				EQUIPMENT DAMAGE			
EQUIPMENT ID	DESCRIPTION	FUNCTIONAL FAILURE MECHANISMS	MTBF (yrs.)	Likelihood	Consequences	Risk	Note	Likelihood	Consequences	Risk	Notes	Likelihood	Consequences	Risk	Notes	Likelihood	Consequences	Risk	Notes
BN01	Bin 1	Structural collapse				5				0								0	
BN01	Bin 1	Corrosion/Wear																	
BN01	Bin 1	Impact damage																	
BN01	Bin 1	Clamshell failure - including actuators																	
BN02	Bin 2	Structural collapse				5				0								0	
BN02	Bin 2	Corrosion/Wear																	

BN02	Bin 2	Impact damage																	
BN02	Bin 2	Clamshell failure - including actuators																	
BN03	Bin 3	Structural collapse				5				0								0	

It's important when considering multiple criterion that everyone understands that the 'Likelihood' being considered is not that of the failure mechanism occurring but rather the Likelihood of the failure having occurred, affecting the criterion of interest. This is an important distinction because the 'failure rate' of the failure mechanism being considered (i.e. the inverse of the MTBF) is common across all criteria while the Likelihood is criteria specific, at least in common cases found in industry.

For example, if we are considering Safety, Environment, Production loss and Secondary damage as our criterion of interest as shown in Table 1 and the failure mechanism being considered is bearing failure of an electric motor, then the bearing failure rate is constant (or approximately so) for the bearing in its operating context while the impact of the bearing failure on Production loss may be major while the same failure may have no impact on the Environment. Clearly from this and other easily demonstrated examples the Likelihood of interest in Criticality analysis is the Likelihood of the failure impacting on the Criteria rather than the Likelihood of the component failing by the failure mechanism of interest.

The 'Risk' associated with a particular failure mechanism and for a specific criteria is calculated by multiplying the Likelihood (L) of the failure impacting on the criteria and the Consequences (C) of failure for that criteria as shown in equation 1.

$$\text{Risk} = \text{Likelihood} \times \text{Consequences} \qquad (1)$$

Equation 1 represents a very simple example of the Risk equation and in some applications many more variables are included [2]. It is important not to make the assessment equation unnecessarily complex as the addition of each new variable increases the complexity of the assessment and obtaining appropriate data to 'feed' the equation can become very difficult. For example, data on the 'Coupling Coefficients' necessary to relate the different criterion is rarely available with any confidence and it is questionable if criteria coupling is logical or desirable in many instances. Similarly combining the MTBF with each criterion is problematic if we cannot allow any instance of a particular failure mechanism and hence will never (hopefully) have enough data to generate a MTBF.

A more subtle problem can be created by making the Risk equation overly complex in that the data used may be of poor quality but may be assumed to be accurate by those using the assessment output. In addition, as our knowledge of the relationships between variables and especially between criteria is often poor reliable data may not be obtainable at the time of the assessment and may not be constant over time.

Table 2

Example of a Likelihood prompt sheet used in a Semi-Quantitative assessment of the Criticality of equipment.

	PRODUCTION	HEALTH & SAFETY	ENVIRONMENTAL	SECONDARY DAMAGE
5	Loss of production is a common or repeating occurrence. Loss of production will result from every failure incident.	Constant. Happens regularly/repeatedly, is expected to happen.	Constant. Happens regularly/repeatedly, is expected to happen.	Equipment damage is a common or repeating occurrence. Equipment damage will result from every failure incident.
4	Loss of production is known to occur, 'It has happened'. Know to occur at other sites within the organization.	Frequent. Has happened, known to have occurred, would not be unusual.	Frequent. Has happened, known to have occurred, would not be unusual.	Equipment damage is known to occur, 'it has happened'. Know to occur in other organizations within the same or similar industries.
3	Loss of production could occur; 'I've heard of it happening' on this or other sites. Know to occur in other organizations within the same or similar industries.	Infrequent. May happen, 'heard of it happening', 'nearly happened'.	Infrequent. May happen, 'heard of it happening', 'nearly happened'.	Equipment damage could occur; 'I've heard of it happening' on this or other sites. Know to occur at other sites within the organization.
2	Loss of production is not likely to occur. Loss of production is considered unlikely even when a failure occurs.	Very Remote. May conceivably happen – but has not happened yet.	Very Remote. May conceivably happen – but has not happened yet.	Equipment damage is not likely to occur. Secondary damage is considered unlikely even when a failure occurs.
1	Loss of production is practically impossible.	Practically Impossible. Difficult to conceive – has never happened in spite of many years exposure	Practically Impossible. Difficult to conceive – has never happened in spite of many years exposure	Equipment damage is practically impossible.

Prompt sheets are vital to assist those carrying out the assessment to make consistent decisions regarding Likelihood and Consequence. The prompt sheets should be completed and agreed before the assessments get underway and training provided to the assessment team members in how to use and interpret these sheets. The facilitator in particular must be very clear on how to interpret these sheets and how they are to be implemented as many of the teams will have questions as to how to apply the prompt sheets when assessing one failure mechanism or other. Table 2 shows a basic Likelihood prompt sheet used in a Semi-Quantitative assessment of Criticality for a major quarrying company. The level of detail included in the prompt sheets will depend on the level of sophistication required and the intended application for the assessment output. In addition, the table should be designed with the available data in mind as there is no value in asking for input that is not available and may only serve to tempt those doing the assessment to 'guess' outside the range of their expertise in an attempt to complete the assessment.

Table 3 shows a basic Consequence prompt sheet used in a Semi-Quantitative assessment of Criticality for the same quarrying company as for Table 2.

Table 3

Example of a Criticality prompt sheet used in a Semi-Quantitative assessment of the Criticality of equipment.

	PRODUCTION	HEALTH & SAFETY	ENVIRONMENTAL	SECONDARY DAMAGE
5	Site wide shutdown, total loss of production for more than 24 hours. NB: downtime could be up to a week.	Major. Fatality.	Major. Extensive environmental damage, damage to company's reputation, major media interest.	Site wide shutdown, multiple unit failure, financial cost >$1M
4	Site wide shutdown, total loss of production for up to 24 hours, or reduced capacity for more than one day.	Serious. Permanently disabling injury.	Serious. Significant environmental damage, local media interest.	Multiple unit failure or shutdown, financial cost <$0.5M
3	Site wide shutdown, total loss of production for up to 8 hours, or reduced capacity for up to one day.	Moderate. Serious, medically treated injury.	Moderate. Moderate environmental damage, no media interest.	Unit shutdown or damage to equipment, some critical equipment failure, financial cost <$100K
2	Equipment failure or shutdown, equipment off-line for up to 4 hours, or reduced capacity for up to one shift. NB: refers to failure of a single machine that does not result in site wide shutdown.	Minor. First aid injury.	Minor. Minor environmental damage.	Some equipment damage, greater than normal repair time (e.g. 1 week), financial cost <$25K
1	Equipment failure that can be repaired on-line and no impact on production or can be repaired on a scheduled maintenance day.	Insignificant. No injury.	Insignificant. Negligible environmental damage.	Minor equipment damage, normal repair time expected, financial cost <$5K

Table 4

Matrix relating Likelihood and Consequences to form a Risk score.

		CONSEQUENCES				
		5	**4**	**3**	**2**	**1**
LIKELIHOOD		25	20	15	10	5
		20	16	12	8	4
		15	12	9	6	3
		10	8	6	4	2
		5	4	3	2	1

'A' - High Criticality

'B' - Medium Criticality

'C' - Low Criticality

5 THRESHOLDS

The raw scores generated by the Criticality assessment are often 'mapped' onto a simpler scale for convenience and it is this criticality score that is used for day-to-day applications. The threshold at which different maintenance or other responses are applied can be determined by nominating risk levels at which some specific activity is considered justified. This threshold risk level can arise from several combinations of the Likelihood and Consequence components. The Criticality derived from the calculated risk or the mapped score is used in the CMMS which is the score used for planning and scheduling decisions.

Table 4 shows the range of raw scores that result from combinations of Likelihood and Consequence scales with the shaded areas representing the thresholds agreed by the participants. The job of establishing the Thresholds should be left to competent staff having experience of the plant and responsibility for the outcomes if things should go wrong. This is the only way to get a serious assessment as well as acceptance of any decisions based on the criticalities.

Threshold levels can be difficult to determine objectively and can lead to many arguments as to the appropriateness of one level over another. For this reason it is helpful to settle any argument before commencing the assessments if possible however as the assessments are carried out using the raw scores the threshold levels can be recalculated later if necessary.

6 SUPPORT INFORMATION

A significant amount of support information is needed to carry out the assessments and as much as possible should be collected and collated before the assessments get underway. Support information required for the criticality assessment includes the following:

- Cost of downtime.
- Labour cost.
- Production schedule.
- Documentation.
- Drawings.

In many instances the information available is uncertain or variable or both, and the use of agreed, representative values is likely to be the most practical way to manage some of this data. For example, the cost of downtime can vary depending on the product being made, the amount of stock on hand and other factors. Similarly labour cost can vary depending on whether the work is carried out on ordinary time or overtime and whether the staff were on site at the time of the failure or were called in.

The availability of relevant documentation and drawings can impact on the analysis as those performing the analysis need to know the interconnection between plant items as well as whether sections of plant can be isolated for repairs or the whole plant has to be taken off-line to make repairs. It is not safe to assume that local staff knows this information and indeed criticality assessments carried out for a range of clients over several years has highlighted significant knowledge gaps among local staff. In particular, operations staff generally has a poor understanding of the maintenance requirements of equipment especially regarding statutory maintenance tasks or overhaul tasks when the link between these activities and daily operations is not immediately visible. Similarly, maintenance staff often fails to fully appreciate the impact certain faults or deficiencies have on plant operability and often do not give the priority to tasks they deserve because of their perception of the importance of these tasks.

The criticality assessment is often the first time that operations and maintenance staff have sat down in 'cross-functional teams' to discuss a range of equipment and its impact on operations. It is common during these assessment meetings to see a growing awareness of the subtleties of the equipment and how these details and subtleties can impact on others.

7 APPLICATIONS

The criticality assessment will be beneficial to the organisation if it impacts positively on decision making. In deciding how best to respond to various situations the criticality ranking is a useful guide in establishing typical responses for various end-user scenarios. These response guidelines can be used to promote uniformity of response by different people across sites and between shifts and provide guidance to those responsible for making decisions in often stressful situations.

When combined with Priority codes, Criticality codes allow automated work scheduling by the CMMS as well as highlighting jobs by their relative importance. Decisions regarding after hours call-in's, break-in decisions, expediting spares and services etc. should be made on the basis of consideration of the combination of Priority and Criticality associated with each job.

The application of the criticality rankings includes the following:

- Prioritization of routine work lists.
- Protecting the schedule.
- Backlog management.
- Prioritization of emergency repairs.
- Guide on effective call-in policies.
- Optimize spare parts holding.
- Review of documents, spares, maintenance plans, etc.
- Support 'Training Needs Analysis'.
- Shutdowns & Turnarounds.
- Development of budgets.

8 PROJECT PLANNING

The development of project plans needs to take account of the criticality of different equipment in the proposed operating system if maximum advantage is to be gained from the plant at the lowest cost. A clear and detailed understanding of equipment and system criticality is vital for decisions on reliability and the need for redundant equipment to achieve the design system reliability. Unfortunately, criticality information is often scarce at the project planning stage and in the absence of a formal assessment process only anecdotal rankings are possible. This is especially so for green field sites or when installing unfamiliar plant in existing sites. It is important to consider the criticality assessment before any major equipment or layout decisions are locked in if possible as it can be difficult and expensive to make changes to the project once underway – usually attracting 'variation' penalties.

9 PLANT DESIGN & REDUNDANCY

One of the most useful outputs from the criticality assessment is the ability to identify where the 'system reliability' is critical and therefore should be subject to additional reliability studies. For example, when the system reliability requirements are greater than can be achieved by one unit alone it may be necessary to install a stand-by unit to ensure the system reliability targets can be met. In most industrial applications the installation of a single stand-by unit is sufficient to meet the system reliability however, in critical industries such as aviation or nuclear power, triple redundancy may be needed. The decision to install additional capacity to protect system reliability will entail significant cost to the organisation and therefore must be based on sound judgement and auditable assessment procedures.

The positioning of installed equipment must also be considered carefully as the trade-off between the need to minimise the footprint of the plant while providing adequate space between the equipment for access and isolation (e.g. in a fire etc.). The criticality assessment needs to consider access and other issues as well as the likely failure mechanisms as the Life Cycle Cost of the plant can be adversely impacted by poor equipment layout and restricted access. It is not uncommon to see equipment sited such that one machine has to be removed to get access to another. More commonly, pipe and cable runs can significantly restrict crane access around machines and maintenance repair times can be extended by these restriction.

10 OPERATING STRATEGY

The manner in which the equipment is used during normal operations has a significant impact on the reliability that can be achieved from individual machines and therefore 'best practice' operating practices must be enforced to maximise the utilisation of equipment and achieve the 'Maximum Sustainable Capacity' (MSC) of the equipment (see Figure 1). Correctly specified equipment that is well operated and maintained has an 'Inherent Reliability' which sets the maximum reliability that can be reasonably expected from the equipment over time. Poor operating practice, inadequate maintenance, inferior parts and materials and poor application will all lead to a reduced level of reliability which will be reflected in the equipment performance over time. If the equipment 'Inherent Reliability' is inadequate for the system then redundancy will be needed if the system reliability is to be achieved.

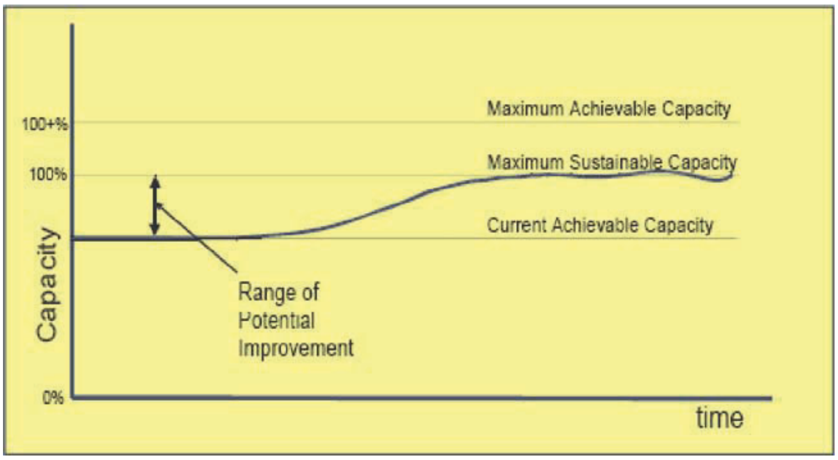


Figure 1 Showing relationship of MSC to MAC

The criticality assessment highlights weaknesses in equipment specification, operation and maintenance which can in turn lead to reduced reliability (compared to the inherent reliability for that item in its operating context).

Stand-by equipment (either active or passive) must be managed in such a way as to maximise the reliability of the system protected by the stand-by unit. One of the most important management considerations therefore is the 'condition' of the stand-by unit at the time it is required. Traditionally, operating practice in many plants has been to operate both the duty and stand-by units for approximately equal times by alternating the units in operation every week or so.

A little thought as to why the stand-by unit was installed should raise concerns about this practice. It is important to appreciate that the stand-by unit was installed to provide an alternative 'function path' should the operating unit fail in service.

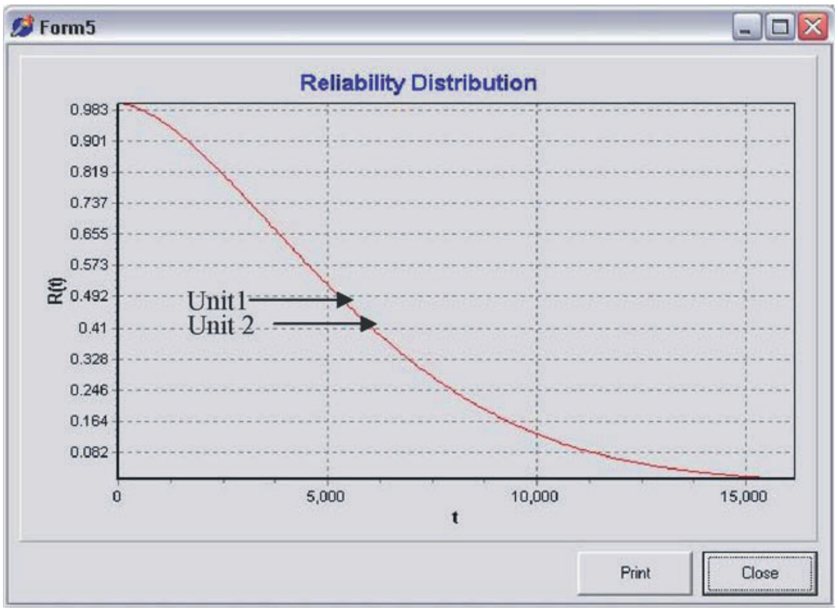


Figure 2 Showing the Duty and Stand-by units in approximately the same condition at the time the Duty unit failed. This is the expected result of operating both units for approximately equal times.

If we accept this logic it is clear that the best way to protect the system is to have the stand-by unit in the best possible condition at all times as this provides the maximum reliability for the unit when it is required (with the possible exception of degrading failure mechanisms unique to the stand-by condition).

Operating both units alternately for approximately equal times will degrade both units which means that when the duty unit fails the stand-by unit is also likely to be in the same condition (see Figure 2).

An alternative operating strategy is to designate one unit as the 'Duty' unit and the second (and possibly subsequent units) as stand-by units. To affect this strategy one unit is used for operations at all times except for when it is in a failed state, at which time the stand-by unit is operated. The stand-by unit under this strategy is in 'good' condition and better able to meet the system requirements while the duty unit is being repaired as indicated by Figure 3.

Repairs to the 'Duty' unit are carried out 'promptly' to return it to service as soon as practicable. It is at this point that the 'Priority' (i.e. how urgently do we need to attend to this task?') becomes important as it is clearly the case that high criticality equipment can have a range of tasks performed at different times some of which require immediate attention while others can be done when an opportunity presents itself. In this way the combination of Priority and Criticality give us an effective tool for decision making and managing a range of assets and systems some of which will be high criticality assets.

In practice, the criticality of the duty unit is usually lower than for the stand-by unit (assuming the 'switch over' time is acceptable) as the system is protected by the stand-by unit when the duty unit fails. This is not the case should the stand-by unit fail as it is only in service when the duty unit is unavailable and therefore failure of the stand-by unit results is system shutdown.

11 MAINTENANCE STRATEGIES

It should be clear from the foregoing discussion that it is the high criticality equipment in particular for which we will be trying to maximise the reliability and therefore every attempt should be made to achieve the 'Inherent reliability' at all times. This will require extensive training for operators and maintenance personnel as well as detailed consideration of the equipment specifications at design time. Poor operating practices can accelerate the degradation of equipment and result in a higher failure rate than necessary. No amount of maintenance will achieve the required reliability if the equipment is inadequate for the application or if it is poorly operated and maintained.

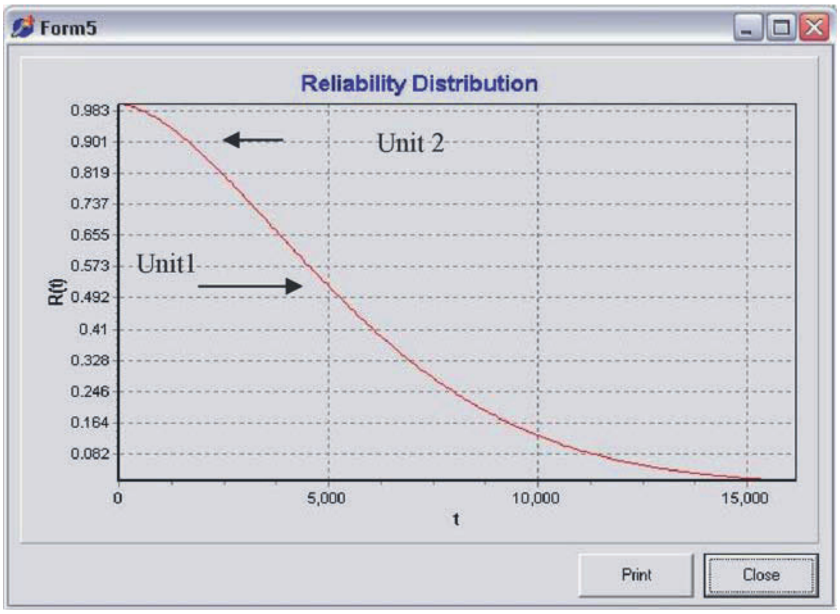


Figure 3 Showing the Duty and Stand-by units at the time of failure with the duty unit reliability reduced and the Stand-by unit still in good condition.

As a general principle every asset should have a maintenance strategy assigned to it, even if that strategy is to 'Run-to-Failure' as in this way at least, the strategy will be the result of deliberate intent rather than as the natural consequence of doing nothing.

Once the maintenance strategies are in place (at whatever level of detail suits the organisation) it is imperative that these strategies are put into effect according to the guidelines of the strategy. If a scheduled replacement is to be carried out at set intervals then these replacements must be done and done on time if the strategy is to be effective. Similarly, if an inspection based strategy is recommended (Condition Monitoring or Non Destructive Inspection or similar) then the scheduled inspections must be done and equally must be acted upon in the event of adverse findings. Failure to implement the agreed strategy will result in an increase in risk of in-service failure and exposure to the resulting consequences of that failure. The criticality assessment documents the consequences of each specific failure considered and as the maintenance strategy is (or should be) based on the findings of the criticality assessment the increased risk associated with failure to implement the maintenance strategy will be high for critical equipment.

It is too often the case that decisions to postpone or delete scheduled maintenance activities are made for narrow, short-term operational reasons (which might be valid in isolation) without a clear understanding or even discussion about the increased risk being assumed by the decision. The criticality ranking allows us to develop agreed policies on how to manage these decisions. Planning and scheduling of day-to-day maintenance activities is one area that regularly highlights this problem.

12 PLANNING AND SCHEDULING

In the words of Peter Drucker "Doing the right thing seems to be more important than doing things right". The relationship between Priority and Criticality can be used as a basis for decision making. By providing planners and supervisors with appropriately developed criticality and priority ratings the company's highest priority tasks will be addressed and addressed first, and decisions on job scheduling and resourcing will be more objective. The Planners can 'push back' against unreasonable requests for 'break-in' work which in the past may have been difficult to resist especially if the request came from a senior staff member. In addition, scheduling and resourcing decisions made by the planners are less likely to be influenced by the pressure applied by those wanting their job addressed first. It is important to establish systems that ensure that it is the company's priorities that are addressed rather than the priority of an individual or area as it is unlikely that addressing the priority of an individual or an area will always be consistent with the organisations priority as a whole.

The aim of Criticality (and Priority) as an Asset management tool is to:

- Use Priority and Criticality to schedule work.
- High Priority/High Criticality work should be done first.
- Provide a communication tool for the organisation.
- Synchronise Operations and Maintenance actions.
- Provide a guide for personnel responsible for managing maintenance work.
- Control set-up, sequencing and completion of maintenance jobs.
- Ensure high priority jobs get done and done first.
- Contribute to the maintenance audit trail from the raising of work through completion and documentation.
- Optimise the utilisation of maintenance labour, both in-house and external.
- Help manage/control costs.
- Help deliver the most cost effective maintenance service possible within the prevailing constraints.
- Help highlight areas where cost effective improvements could be made.
- Minimise downtime of critical equipment.
- Achieve 'Maximum Sustainable Capacity' of critical equipment.
- Assist in achieving zero Lost Time Injuries and zero Medical Treatment Injuries.
- Maximise the usefulness of spare parts held.

Once the maintenance schedule has been generated using a combination of Priority and Criticality it is important to protect the schedule from undesirable last minute changes. The task of protecting the schedule is complicated by the fact that competing aims are in play, namely:

- Resist changes to the schedule to ensure that planned jobs can go ahead as planned.
- Allow changes to the schedule where the change is in the best interests of the organization.

The key to finding the correct balance between these competing aims is to ensure that the jobs on the schedule are those with the highest combination of Priority and Criticality. Any late addition or changes to the schedule will therefore have to have a higher combination of Priority and Criticality than the lowest ranked job on the schedule before it can displace one or more jobs off of the schedule. This approach has the virtue of resisting unjustified changes while allowing those changes that are in the interest of the organization.

Without a clear set of guidelines by which the planners and supervisors can agree what is in the upcoming schedule, the decision will be made by whoever has the loudest voice or the most influence rather than in the best interest of the organization. Failure to provide 'tools' such as criticality rankings and priority rankings for use by operators, maintenance and supervisory staff leads to arguments as to which jobs should be done and in what order and quickly undoes any attempt to introduce planning and scheduling to the organization.

13 BACKLOG MANAGEMENT

Backlog management quickly gets highlighted whenever maintenance and operating practices are being reviewed and is a common point of conflict between the operations and maintenance groups. Much of the conflict arises because of a poor understanding of what jobs should be in the backlog and which shouldn't. It is common practice in many organizations to consider any job known about but not yet completed, to be included in the backlog. This approach results in a large backlog with some jobs remaining there for weeks, months and even years, in some cases only to be deleted later.

A better approach is to define backlog as being known jobs not completed on time or having very little chance of being completed on time. Some of these jobs cannot be completed on time while others may be completed with additional resources such as overtime, contractors etc. All known jobs which are not overdue and which have a high likelihood of being completed on time are not backlog jobs and can be defined as forward log if one wishes. The forward log therefore is made up of known jobs that need no additional efforts over and above normal practice to be completed.

It is sobering to realize that if the 'all jobs' definition of backlog is retained as the organization moves towards planned maintenance then the backlog will approach infinite size as the amount of jobs in the backlog would include all future scheduled jobs up to the end of the chosen planning horizon.

Using the recommended definition of backlog mentioned above (… known jobs not completed on time or having very little chance of being completed on time) a basic set of rules can be agreed by the operations, maintenance and management to ensure that the total job count or man-hours backlog remains low and contains only low criticality jobs. An example of rules agreed by one site is shown here.

- No High Criticality equipment in the backlog.
- Less than 10% of backlog to be Medium Criticality equipment.
- Rest of backlog is Low Criticality equipment.

These or similar rules can be established to avoid conflict and ensure that the important work is done and done first. If the amount of high priority work exceeds the available capacity it is unacceptable to reprioritize some of the work to lower levels just to keep within resource constraints and additional resources must be found or the risk associated with not completing the job on time managed some other way (e.g. may have to shut part of the plant in extreme cases).

14 SPARE PARTS MANAGEMENT

The reason for keeping an inventory of spare parts is to control or avoid the consequences of failure of the equipment. Spare parts have to be managed to maintain the optimum trade-off between the cost of purchasing/holding spare parts and the consequences of NOT having these spares available when required. When the cost of sourcing a spare part off-site only when required (considering all costs especially downtime and lost product) is significantly greater than the cost of holding the spare in-house, the spare is usually held on site. This superficially simple cost-benefit analysis can be very difficult to carry out in practice especially for those parts infrequently or irregularly used or if their impact cannot be easily quantified in dollar terms.

Over time spare parts holdings can increase to unacceptably high levels due to the generally conservative nature of maintenance personnel and the difficulty in removing redundant and overstocked spares from the store. It is common practice in many organisations to charge all spares to the maintenance budget when booked out of the store which in the case of redundant spares is a cost without a benefit to the maintenance manager and therefore the maintenance department is more likely to leave the spare in the store rather than absorb the cost of booking it out if there is no use for that spare.

Over stocked spares are usually easier to handle as the spares holdings can be run down over time and a lower maximum stock level set for future purchases. When control of the store is in the hands of the Administration department rather than the Engineering/Maintenance department there is little incentive for the maintenance staff to pursue spares reduction/optimisation.

Notwithstanding the difficulties mentioned above there are significant opportunities to reduce the spare parts holding of an organisation without increasing the overall risk to the business to unacceptable levels due to not holding adequate spares. ***This is not to say that reducing spares holdings while accepting a controlled increase in risk is not a viable business strategy***. The way to achieve this gain is to ensure that the parts held are biased towards the high risk assets and that sufficient numbers of these spares are available when needed. The optimisation of inventory is achieved by balancing the losses due to stock-outs or delays with the cost of holding the parts until needed. In addition, savings can be achieved by eliminating parts held for low to very low risk assets. In addition to the reduction in stock levels and the cost savings arising from that reduction the spares held are now known to be aligned to the high risk assets on the site and the levels held are known to be optimum or near optimum for the current operating context. The key to achieving the potential savings is in knowing where the asset related risks are in your business and tailoring the spare parts holdings to meet the needs of these high risk assets. The criticality assessment provides a basis for spare parts analysis by identifying ALL the assets in the area of interest, their components, parts and failure modes. In addition, the failure rate, cost of failure, downtime cost, labour cost, repair cost, safety & environmental implications etc. are generated for each failure mode.

15 DOCUMENT MANAGEMENT

Documents such as operating manuals, maintenance manuals, drawings etc. are a vital resource for the organisation and must be the subject of rational management strategies just as hardware or other assets might be. Documents can be difficult and expensive to obtain and in the case of drawings expensive to produce. Over time, drawings in particular will become out of date as changes are made to the plant requiring the relevant documentation to be updated. Frequently the resources are not available for this vital task and over time the documents become outdated and increasingly inaccurate. The safety implications of incorrect or inadequate data in particular means that documents associated with critical equipment must be managed well and kept up to date.

The criticality assessment and ranking can provide a guide for periodic review and updating of documentation associated with high critical assets as a minimum and a document management policy should be developed to manage these documents over time. Electrical drawings and pressure and steam system drawings if incorrect can lead to injury or fatality and only by ensuring that relevant documentation is regularly reviewed in line with its criticality can this danger be minimised.

16 BUDGET MANAGEMENT USING CRITICALITY

Across the board reductions in budgets have become almost routine in many organisations in pursuit of cost reductions and ever better efficiencies. Most common is the percentage based reduction over a previous budget (e.g. 10% reduction in the maintenance budget compared to the previous year) which is often repeated in subsequent years and by subsequent managers. In many instances these budget cuts are well reasoned and can be beneficial if well targeted and applied. In addition, tightening of resources can trigger some degree of innovation amongst staff however it is important not to get carried away with this logic as the extreme of too little resources will inevitable result in some loss of function whether intended or not.

Too often this policy is put into effect by the elimination of previously favoured programs especially those involving inspections or other proactive tasks as they are seen not to provide immediate effects (especially by operations staff) and may have some level of inconvenience associated with their implementation. In the short to medium term therefore the elimination of these programs can create the illusion of reducing costs and streamlining operations. It is however often the case that when these programs are eliminated that their impact cuts across all equipment effected by the program regardless of criticality.

Clearly a better approach would be to reduce activities on low criticality equipment and systems in a targeted way until the cumulative savings achieved in this way meet the required budget reduction. In this way the manager knows that the cuts have been applied to the lowest criticality equipment (or activities) and the likely increased risk traded off against the cost savings, are clearly identifiable and quantifiable.

17 CONCLUSIONS

Criticality and its close relative Priority are significant components in many operations and maintenance functions underpinning decisions regarding equipment operating strategies, maintenance strategy development, planning and scheduling of routine and ad-hoc jobs and backlog management. In addition, the criticality ranking assists in decision making regarding project design, redundancy, spare parts, training, document management etc. and therefore should be carefully developed and kept up to date over time.

If those charged with implementing the operating and maintenance functions of the organisation are not provided with a structured, solidly based criticality ranking of the equipment and systems in their care individuals will attempt to assign criticality (often confusing it with priority) in an attempt to get some deserving job completed. For this reason it is important that the organisation develops a clear indication of the criticality of its assets using a systematic approach which is sufficiently detailed, rigorous, auditable, and reviewable and can be updated over time. Of major importance it the need for widespread acceptance of the criticality by those who will have to work with it and therefore wherever possible, a wide range of people should be allowed to make contributions to the assessment consistent with the need to do so efficiently, effectively and cheaply.

18 REFERENCES

[1] Standards Australia (1999), Risk Management AS/NZS 4360:1999, Standards Australia Publications.

[2] Engineering Specifications and Standards Department (1980), Procedure for Performing a Failure Mode, Effects and Criticality Analysis, Military Standard 1629A, US Department of Defence Publication.

19 DEFINITIONS

Criticality – **Static assessment of equipment or system importance**. Criticality is a fixed value that depends on the role of the equipment in the business and only changes when the operation changes.

Priority – **Dynamic assessment of activity importance**. Priority varies for each task to be carried out on equipment regardless of the criticality of that equipment.

MSC – 'Maximum Sustainable Capacity'. The equipment capacity that can be achieved in a sustained way over an indefinite period of time.

MAC – **'Maximum Achievable Capacity'**. The maximum capacity that can be achieved from the equipment over a short period and usually results in significant degradation of the equipment and therefore is not sustainable.

CAC – 'Current Achievable Capacity'. The currently achieved steady state capacity.

Backlog – 'A list of jobs requiring extraordinary efforts to complete'**.**

Asset – 'Hardware, software, procedure etc. used to provide a valued function'**.**

A REVIEW OF THE MIMOSA OSA-EAI DATABASE FOR CONDITION MONITORING SYSTEMS

Avin Mathew, Liqun Zhang, Sheng Zhang, and Lin Ma

Cooperative Research Centre for Integrated Engineering Asset Management, Queensland University of Technology, George Street, Brisbane, Queensland 4001, Australia.

Abstract: Condition monitoring systems are prevalent in industries that contain critical assets. Their use allows the detection of potential failures at an early stage in order to minimise downtime and maintenance costs of a system. Many condition monitoring systems are commercially available, however these systems are limited in their support of diagnostic and prognostic models developed by external parties. An attempt to develop an extensible condition monitoring software lead to the creation of an ISO 13374/OSA-CBM based system named BUDS, which focuses on vibration condition monitoring. Research into condition monitoring data management strategies led our team to implement a database to support the data-driven application. A database based on the MIMOSA OSA-EAI database specification was used to store asset and sensor location, measurement event, raw sensor and processed signal data. This paper investigates the use of the MIMOSA OSA-EAI database for condition monitoring systems, and presents several issues encountered during the development process.

Key Words: Condition Monitoring, Critical Assets, Downtime, Diagnostic, Prognostic, BUDS, MIMOSA, Sensor, Development.

1 INTRODUCTION

Condition monitoring systems have become more popular in recent years due to lower financial barriers to entry. Sensor technology is constantly improving in detecting more types of conditions and analysis models are becoming more mature. In an effort to expand the limits of traditional commercial condition monitoring systems, we developed a condition monitoring system, BUDS, which supports advanced diagnosis and prognosis models not available in commercial systems. Based on the Open Systems Architecture for Condition Based Maintenance (OSA-CBM), BUDS comprises of a single component of a larger asset management decision support system that also implements reliability and decision making modelling.

A major component of condition monitoring systems is data management. Although vast amounts of measurement and analysis data flow through condition monitoring systems, data management is not usually given high priority due to its less visible nature from a user perspective. Using a standards-based approach, BUDS implements a MIMOSA (Machinery Information Management Open Systems Alliance) OSA-EAI (Open System Architecture for Enterprise Application Integration) derived database. The MIMOSA OSA-EAI is a standard for data exchange of engineering asset management data. A detailed summary of the OSA-EAI can be found in Section 3.

In this paper we highlight the issues we faced in using the database specification from the MIMOSA OSA-EAI for BUDS. The issues presented are from a development viewpoint, and aim to assist the development of future condition monitoring systems based on the OSA-EAI.

2 BUDS

BUDS is based on the ISO 13374/OSA-CBM model of development. ISO 13374 is a two part standard that publishes general guidelines on the architecture of condition monitoring systems. It presents a modular approach to system design in order to allow for interoperability between various condition monitoring components and external systems. The standard was largely derived from the Open Systems Architecture for Condition Based Maintenance (OSA-CBM). Both standards comprise of a six block structure involving data acquisition, data manipulation, state detection, health assessment, prognostic assessment, and advisory generation. While the ISO 13374 standard provides guidelines on how to develop a condition monitoring system, the OSA-CBM provides the addition of tools for implementation.

There have been several implementations of the OSA-CBM. A vehicle health management system based on the OSA-CBM was developed by Boeing [2, 3]. Another OSA-CBM system was used in the diagnostic and prognostics for avionics health management [4]. Another implementation was used to monitor an electro-hydraulic test rig [5]. A team at Penn State University used the OSA-CBM for a generic monitoring application [6]. The data management strategies employed by each of the aforementioned projects are unclear.

BUDS is being developed as part of a larger asset management software suite that also supports reliability and decision making models. The condition monitoring component is essentially a module within the larger asset management system. Due to the extensible architecture of the system, different signal processing models can be exchanged depending on the monitoring situation. An integrated database supports the data requirements of the system. The MIMOSA OSA-EAI was selected as a design for the database for several reasons:

- The OSA-CBM is based on sections of the OSA-EAI specification, providing for easier integration;
- The specification has been designed and validated by asset management experts;
- Following a standard specification will provide greater interoperability with other systems.

The merits between a database and a file system for condition monitoring systems have been debated before [7]. A database is not a panacea, and the justification of the data management strategy employed must be based on the functionality of the condition monitoring system. The data used by BUDS is largely relational. Asset registry information, such as asset types and model numbers, need to be linked to current and historical locations of the asset, measurement points on the asset, sensors installed at the measurement points, readings taken from sensors, and calculations performed on those readings. The housing of such relational information is best performed through use of a relational database. Apart from the numerous advantages in storing information in a database, additional advantages can come through possibilities for interoperability of programs [8]. As database programming interfaces are open and well documented, a potential central data repository can be created more easily [9]. Given the provision that the programs follow the same database specification – in this case, the MIMOSA CRIS – multiple programs can access information within the same database. This leads to a reduction of redundant data and may streamline data management processes within an organisation.

3 MIMOSA OSA-EAI

With the numerous asset management systems offered by different vendors, the processes of integration can be problematical as many systems have their own unique data exchange interfaces. This leaves businesses facing a dilemma, as different integration techniques bring their own advantages and disadvantages. Purchasing systems from a single vendor leads towards system compatibility, however suppliers may not provide a total asset management solution, and the reliance on one vendor can prove risky. Businesses may purchase a custom bridge that integrates different systems, which may prove more cost effective than internally building one, but provides less customisation ability and requires updates for new system versions. Another option is to use an industry-standard bridge, which allows businesses to mix different systems with reduced integration costs. However, there may be performance loss compared to a custom solution and vendors must be willing to support the standard.

The absence of a standard for asset management data exchange was a driving factor in the formation of MIMOSA and the subsequent development of the OSA-EAI. The OSA-EAI provides open data exchange standards in several key asset management areas: asset register management; work management; diagnostic and prognostic assessment; vibration and sound data; oil, fluid and gas data; thermographic data; and reliability information. These seven areas are defined by a relational model named Common Relational Information Schema (CRIS). The CRIS defines asset management entities, their attributes and associated types, and also relationships between entities.

As seen in Figure 1, a reference data library sits on top of the CRIS. The library contains reference data compiled by MIMOSA which can be stored by the CRIS and are intended to facilitate communication between MIMOSA-compliant systems. The reference data primarily consist of ‘type’ information such as asset, segment, and event types. However, the largest component of the reference library is manufacturer details.

The OSA-EAI package contains SQL (Structured Query Language) scripts for creating a database based on the CRIS and inserting data from the reference library. A program does not need to implement the database component to be MIMOSA-compliant – only the XML schema must be implemented. However, a MIMOSA database implementation makes future development significantly easier in order to comply with the MIMOSA standards.

One key advantage of using the OSA-EAI is through software reuse. Software reuse is a technique that aims to develop components that can be implemented in multiple systems with slight or no modification. Reusable components increase the likelihood that prior testing has been undertaken, and that common bugs have been detected. Another significant advantage in reuse is the potential reduction in implementation time for not only the development stage, but also those of design and testing. As MIMOSA have undertaken a comprehensive analysis of the data elements involved in engineering systems, harnessing their effort avoids duplicating existing research.

Another key advantage of using the OSA-EAI is that of data interoperability. Our experience has found many condition monitoring systems storing information in proprietary file formats; making it difficult to use the data with external programs, such as analysis tools. Unless the software supports an export facility, organisations are effectively locked into particular tools. An open specification for data management would assist in interoperability between engineering asset management systems.

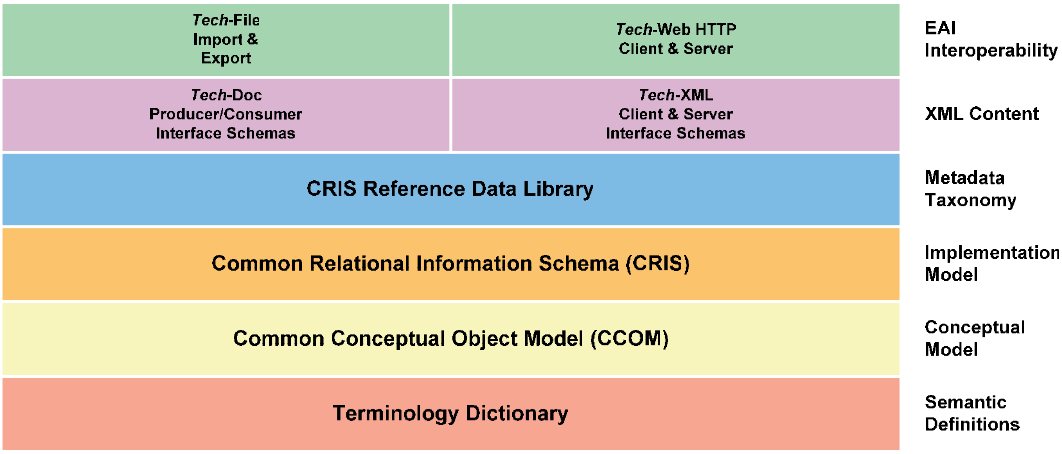


Figure 1: MIMOSA OSA-EAI layers

The suitability of the OSA-EAI for condition monitoring programs stems from its heritage as a standard for vibration data exchange. The current version of the OSA-EAI is markedly different from its predecessors, and has been expanded to support sound monitoring; fluid, air and solid sampling; and thermographic and image monitoring.

Documented implementations of the OSA-EAI specification are rare. There are currently two systems that are certified as meeting the MIMOSA standard, Emerson Process Management RBMware 4.2, and Rockwell Emonitor Odyssey 1.2. Both of these programs implement the OSA-EAI version 1.0. Various other providers including DMSI, Ivara Corporation, Ludeca, SKF, and Vibration Consultants have some compatibility with the OSA-EAI, but are not compliant. Although most projects that use the OSA-EAI inevitably implement the XML communications layers and forego the database element, one exception is a condition monitoring system developed by a Chinese university [1]. The system used an OSA-EAI database, however, the scope of the usage was very narrow. Only three tables were used in their implementation: those of `segment`, `site`, and the `unit_type` tables.

4 MIMOSA OSA-EAI DATABASE ISSUES

During the development of BUDS, several problems were identified in using the OSA-EAI database. Most issues could be resolved with changes to the data model, and these are outlined in Section 4.1. However certain issues were too large to be resolved, such as poor documentation, and these are detailed in Section 4.2.

4.1 Resolvable Issues

Resolvable issues are those where direct management action could be undertaken by our team. These included the customisation of the data model through additions, deletions, and renaming; the modification of reference data; and dealing with design ambiguities.

4.1.1 Data model simplification

The CRIS tries to provide a generic relational model for a wide spectrum of asset management firms and systems. Because of the abstract model on which it is based, many supported features are not needed in a pure condition monitoring system.

Naturally, work management, ordered list, network and function tables will not be used. To tailor the CRIS to meet our needs, a two stage simplification process was conducted.

The first simplification step was to identify unwanted tables in the database. This involved firstly identifying the functions of the system and data required for the database, the results of which filtered the remaining tables. To assist with this process, a tool was developed that would determine the minimum set of tables required to support a particular input table. By examining the non-null foreign key relations, a depth first search is performed to determine the required tables. For example, Figure 2 shows the minimum number of tables required to support a measurement event record. The rounded boxes represent the tables, while the square boxes on the arrows represent the non-null foreign keys.

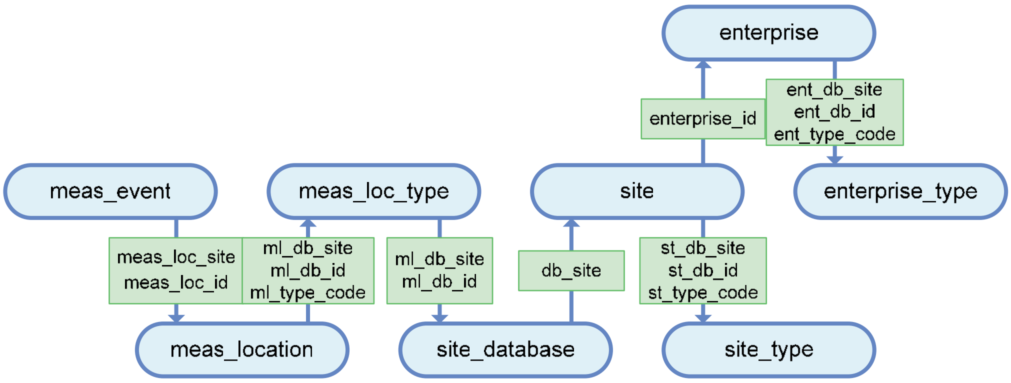


Figure 2. Minimum set of entities to support `meas_event`

The second simplification step was to identify unwanted columns in the data model. The CRIS employs a complex tracking system that distinguishes which database information is derived through the `site_database` table. This leads to many situations where three fields are required to represent a foreign key code. For example, as seen in List 1, the asset table requires three fields (`as_db_site`, `as_db_id`, and `as_type_code`) to define an asset type, three fields (`mf_db_site`, `mf_db_id`, and `mf_duns_code`) to define the manufacturer's DUNS code, and three fields (`asr_db_site`, `asr_db_id`, and `asr_type_code`) to define the asset readiness state.

`<table name="asset">`	
`<column name="asset_id" type="unsignedInt"/>`	
`<column name="as_db_site" type="string"/>`	
`<column name="as_db_id" type="unsignedInt"/>`	Asset type
`<column name="as_type_code" type="unsignedInt"/>`	
`<column name="mf_db_site" type="string" minOccurs="0"/>`	
`<column name="mf_db_id" type="unsignedInt" minOccurs="0"/>`	Manufacturer's DUNS code
`<column name="mf_duns_code" type="unsignedInt" minOccurs="0"/>`	
`<column name="asr_db_site" type="string" minOccurs="0"/>`	
`<column name="asr_db_id" type="unsignedInt" minOccurs="0"/>`	Asset readiness state
`<column name="asr_type_code" type="unsignedInt" minOccurs="0"/>`	
.	
.	
.	

List 1. MIMOSA OSA-EAI `asset` table definition

BUDS is designed to run on a single database architecture. Any intersection of data from multiple BUDS instances should be equivalent, and hence, the advantages of supporting `site_database` information is negated. Simplification involved converting such triple column keys into single columns by removing the `db_site` and `db_id` columns. Consequently, the primary keys in the linked tables were reduced from three to one column. This reduction dramatically reduced the system development overhead for managing tables.

4.1.2 Data model additions

While the OSA-EAI provides a comprehensive data model, several additions were required during the development of the condition monitoring system. These additions were made to support particular functions in the condition monitoring system.

The condition monitoring system provides an automated expert system for the diagnosis of faults. It stores FMEA information such as symptoms, causes, and remedies. To provide tight integration between the additions and the standard MIMOSA data model, symptoms are linked to asset types. To assist with automated analysis, symptoms are recorded with specific types, such as physical symptoms, vibration time domain symptoms, or vibration frequency domain symptoms.

To assist in the administration of asset registry information, another table was added that related numeric data to asset types. As mentioned in Section 5.3, the numeric data tables provided a flexible structure to add numeric data to other tables. A filter in the condition monitoring system limits the displayed numeric data types to only those relevant for a particular asset type. For example, rated speed is suitable characteristic for a motor asset type, however, it is nonsensical for a power transformer asset type.

Because analysis is rarely undertaken on a raw signal, the data must first be processed to calculate relevant features. These features can then be used in diagnosis or prognosis analysis. One of the functions in BUDS is to process raw signals and extract time domain (e.g. kurtosis) and frequency domain (e.g. band frequencies) features. There are two motivating factors on storing these features. Firstly, the features are computationally expensive to calculate, so retrieving a previously stored value would increase system performance. Secondly, the features can act as a form of data compression (albeit lossy) as once the features are calculated from the signal, the original raw signal can be deleted.

Figure 3 illustrates the three typical data structure models used in database design. Methods 1 and 2 hard code features, while Method 3, as followed by MIMOSA, provides extensibility at the cost of performance. As opposed to the flexible structure that MIMOSA generally adopts, BUDS stores the features using Method 1. With vibration data, 20 features are calculated from a time waveform record. Storing these features using Method 3, means 20 rows per time waveform record, as opposed to one row using Method 1. For multi-asset/multi-site monitoring, our calculations showed that the feature table would be receiving a million rows within months of operation, hence the selection of Method 1. Method 2 was not selected as it became unwieldy when dealing with a large number of features.

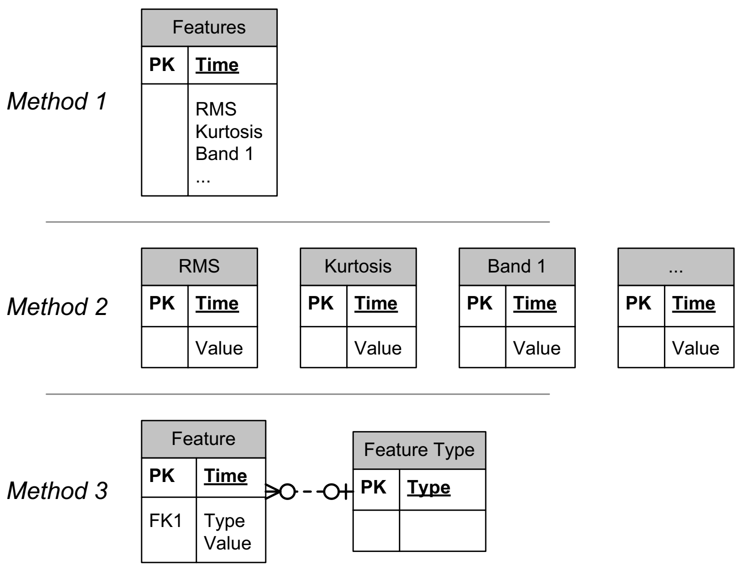


Figure 3: Foreign key ambiguity

4.1.3 Miscellaneous Issues

Hierarchies in type tables — Some type tables in the OSA-EAI support hierarchical data. For example, the `asset_type` table has an associated `asset_type_child` table which records parent and children pairs between different asset types. However, an issue of data redundancy occurs in the reference data supplied by MIMOSA. The duplication of the hierarchical structure can be seen in Listing 2. The `asset_type_child` table stores numerical parent and child identifiers, while the `name` column of `asset_type` stores the hierarchy by separating items by commas.

Table 1 Hierarchical `asset_type` and `asset_type_child` reference data

as_type_code	name
374	Bearing
367	Bearing, Journal
363	Bearing, Journal, Elliptical
365	Bearing, Journal, Plain
366	Bearing, Journal, Pressure Dam

as_type_code	child_as_type_code
374	367
367	363
367	365
367	366

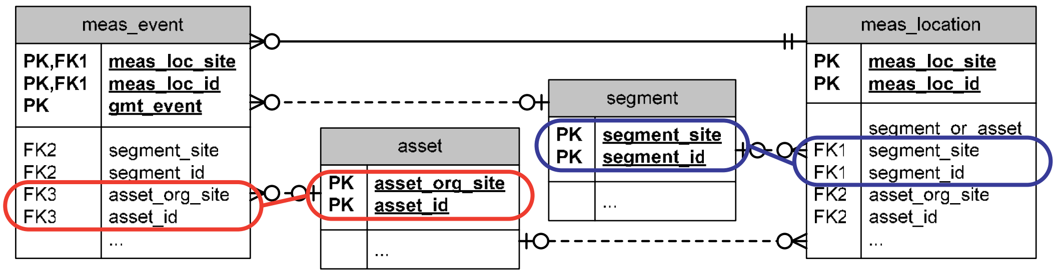


Figure 4: Foreign key ambiguity

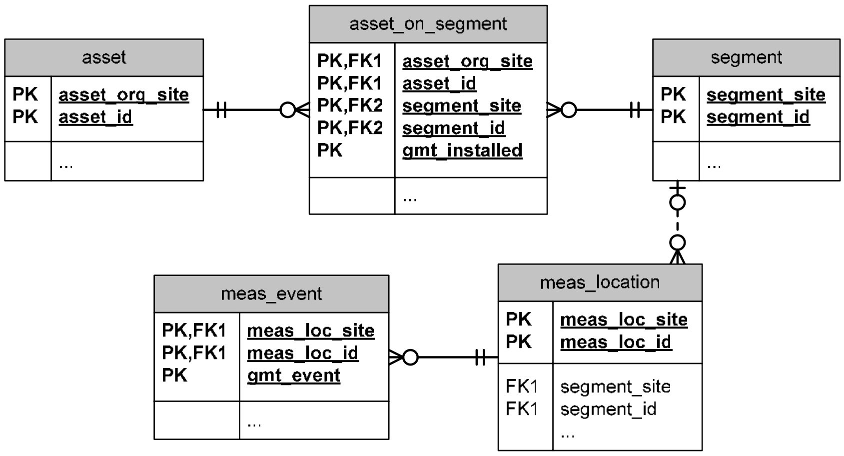


Figure 5: Normalised data model

From both a system development and data modelling perspective, the MIMOSA implementation is not optimal. The commas in `name` make the development of hierarchical views in the condition monitoring system more difficult, because content and presentation are combined. The `name` column cannot be used directly, as it presents a confusing display for users. It must be first processed to remove the parent component – that is, only the element after the last common remains. Normalised data models should not contain duplication of information and the situation should be remedied by removing the parent hierarchy in the `name` column of the `asset_type` table.

Country specific columns — The `manufacturer` table contains two columns `us_ph_number` and `us_fax_number` which store United States phone and fax numbers respectively. Distinguishing between US and international numbers is an unusual inconsistency in the generic nature of the OSA-EAI. While manufacturer contact information is not critical for condition monitoring, the offending columns were removed.

Table and column names — The naming of certain tables and columns is somewhat inconsistent. Abbreviations of a term are used, while in other instances, the full term is used, and in some cases, different abbreviations area used. For example, `meas_event` is abbreviated to `mevent` in the `mevent_chr_data` table, presumably to reduce the length of the name. Consistent naming makes system development easier and hence we adopted a naming scheme that only used one version (either the full name or abbreviated) for table and column names. Thus the `mevent_chr_data` was renamed to `meas_event_chr_data`.

Foreign key ambiguity/Denormalisation — One issue that arises with the `meas_event` and `meas_location` tables is relating measurement events and locations to assets and segments. As Figure 4 shows, both the `meas_event` and `meas_location` tables can be associated with either the `asset` or `segment` tables. The OSA-EAI only provides the constraint on `meas_location` that either the asset or segment but not both should be populated which stands to reason that `meas_event` must undergo the same constraint, lest the data model be denormalised. As measurement locations and segments are very similar conceptual entities, the approach that we viewed as most logical is outlined by the coloured boxes in Figure 4 (`meas_event` with `asset`, and `meas_location` with `segment`).

However, the discussion becomes moot when the `asset_on_segment` table is considered. This table details asset utilisation on a segment, and as it also joins the `asset` and `segment` tables, it indicates a denormalised data model. Normalisation is a goal for proper database design, and a normalised model which implements the decision about joining `meas_location` and `segment` is presented in Figure 5. From the figure, the reasoning behind the inclusion of the extra columns becomes apparent. To find a particular asset on which a measurement event was taken now requires a query that contains four joins, while previously it required just one. While the latter approach is more correct from a data modelling perspective, our database implementation used the former, since the issues associated with a denormalised model would be insignificant.

Linked data limits — To provide flexibility for cross-domain implementations, many core tables are linked to data tables which can be used to dynamically add virtual columns to the core tables. Attributes can either consist of character data, or numeric data. The data tables typically consist of the attribute type, a value for the attribute, and also the unit type of the attribute. While such a model provides flexibility, it results in a slightly longer development time, as instead of hard coding particular columns, a generic approach is required to handle 'virtual' columns.

The OSA-EAI has such data tables for the `asset`, `segment`, and `meas_event` tables. The OSA-EAI places limits on the number of data that can be attached to records in the former tables. Table 2 indicates the restrictions placed on the asset, segment, and measurement event tables. For example, an asset cannot be linked to multiple voltage characteristics and measurement events cannot be linked to two pressure values whose units are Pascals. As there seems to be no justification for this restriction, to overcome this problem, an extra `ordering_seq` (named to follow the OSA-EAI conventions) primary key field that acts as a counter was added to the linking tables (`as_chr_data`, `as_num_data`, `sg_chr_data`, `sg_num_data`, `mevent_blob_data`, `mevent_chr_data`, and `mevent_num_data`).

Table 2 Restrictions on linked data tables

	Type of linked data	Restriction
`asset` / `segment`	Numeric / Character	Unique numeric / character type
	Binary	None
`meas_event`	Numeric / Character	Unique unit type
	Binary	Unique binary type

4.2 Integral Issues

This section details those issues where it was impractical or valueless to modify the specification due to the sheer size or integral nature of the issue. They are included in this discussion to highlight certain problems that may affect future implementations of the OSA-EAI. Two issues are discussed: that of insufficient documentation on the specification, and the database design choices that must followed to retain a consistent data model.

4.2.1 Documentation

The largest issue our team had with the MIMOSA OSA-EAI was the lack of comprehensive documentation for the CRIS. The documentation for the CRIS is currently supplied through a single Microsoft Word document. The format of the documentation is such that each table is allocated its own page containing the name of the table, a description of the table, columns of the table, a description of the column, and the type of the column. Primary key and foreign key information are also included. For main tables, such as site, asset, and measurement event, descriptions adequately describe the function of the table. However, there remain many smaller tables where descriptions are not evident or even non-existent.

Many descriptions for attributes are also non-existent or purposefully omitted. For example, the critical `meas_event` table has no descriptions for ID, site and code columns. While column codes can eventually be recognised through repetitive familiarisation, even simple descriptions would assist in learning for beginners.

Apart from providing more comprehensive textual descriptions for entities and attributes, a suggestion for the CRIS is to restructure the documentation to improve navigation. It is currently difficult to traverse through the Microsoft Word document, particularly with the current formatting where primary and foreign key information is kept separate. Instead of a separate section for keys, columns related to primary and foreign keys should be marked as such in place, and have links to associated table pages. Although the Portable Document Format is a better option than the Microsoft Word format in terms of document portability standards, moving to a more linked structure, similar to web-based documentation, would enhance learning and ease of use.

Additional documentation to select problems can be found on message forums available through Yahoo! Groups. It forms an interactive knowledge base by providing a method of contacting developers and other MIMOSA users for help, where all messages and replies are publicly accessible.

4.2.2 Surrogate keys

From a data modelling perspective, parts of the OSA-EAI is built on surrogate keys rather than natural keys. That is, the primary key for an entity is a numeric value that has no real-world relation to the record it identifies. The advantage of surrogate keys is that changes to primary key values can be performed easier, as the surrogate key remains unchanged, while the data attached to it is changed. Such keys may also improve the performance over situations where the primary key may consist of a multi-column natural key. The disadvantage is that queries generally require more joins, although joins are easier with single rather than multi-columned keys. From a condition monitoring system development point of view, apart from the increased complexity of queries, management of data is made easier using a surrogate key approach.

4.2.3 Excessive normalisation

One of the criticisms voiced by industry about the OSA-EAI is the excessive normalisation that takes place in the data model. While normalised models are recommended for transactional database systems for recording and managing data, they are not as suitable for analysis systems. As condition monitoring systems serve both as a recording and analysis tool, a compromise must be made. As discussed prior, the OSA-EAI attempts to produce a generic model for asset management data, and consequently, complexity is increased. Linking tables for adding characteristic data to entities are included to provide extensibility, but the downside is the increased number of joins for SQL queries. In addition, the OSA-EAI is laden with nearly every attribute being referenced as a foreign key to another table. In the administration of table keys, the surrogate key approach keeps key naming consistent.

5 CONCLUSION

The MIMOSA OSA-EAI is a thorough and well constructed specification for asset management data. The CRIS provides a usable database implementation which can be populated by a comprehensive reference data set. With a few adjustments, the OSA-EAI is suitable for a condition monitoring database, covering the major aspects of condition monitoring, including asset and sensor registry management, measurement event management, and storing raw and processed signals. The primary issues encountered for condition monitoring system development were the lack of documentation and the complexity of the data model. Peculiarities also arose in the reference data and certain sections of the data model. To support advanced trending and diagnostic functionality, additional tables were required. Although the MIMOSA OSA-EAI continues to be a work in progress, it provides a bright future for engineering asset management systems.

6 REFERENCES

[1] Zhang L, Xu F, Jia M, Zhong B & Liu Z. (2002) The design and realization of a portable and permanent integrated condition monitoring system. *China Mechanical Engineering*, 13(10), 860-863.

[2] Followell D, Gilbertson D & Keller K. (2004) Implications of an open system approach to vehicle health management. *Proceedings of the IEEE Aerospace Conference*, Big Sky, Montana, USA. pp.3717-3724.

[3] Keller K, Wiegand D, Swearingen K, Reisig C, Black S, Gillis A & Vandernoot M. (2001) An architecture to implement integrated vehicle health management systems. *Proceedings of the IEEE Systems Readiness Technology Conference*, Valley Forge, Pennslyvania, USA. pp.2-15.

[4] Byington CS, Kalgren PW, Dunkin BK & Donovan BP. (2004) Advanced diagnostic/prognostic reasoning and evidence transformation techniques for improved avionics maintenance. *Proceedings of the IEEE Aerospace Conference*, Big Sky, Montana, United States. pp.11.

[5] Chidambaram B, Gilbertson DD & Keller K. (2005) Condition-based monitoring of an electro-hydraulic system using open software architectures. *Proceedings of the IEEE Aerospace Conference*, Big Sky, Montana, United States. pp.8.

[6] Lebold M, Reichard K & Boylan D. (2003) Utilizing DCOM in an open system architecture framework for machinery monitoring and diagnostics. *Proceedings of the IEEE Aerospace Conference*, Big Sky, Montana, United States. pp.10.

[7] Qi K & He Z. (2005) Data management in monitoring and diagnosis system for electromechanical equipment based on LabVIEW. *Computer Measurement & Control*, 13(1), 24-30.

[8] Inaudi D. (2002) Development of reusable software components for monitoring data management, visualization, and analysis. *Proceedings of the SPIE International Symposium on Smart Structures and Materials*, San Diego, California, USA. pp.10-16.

[9] Butcher SW. (2000) Assessment of condition-based maintenance in the Department of Defense. McLean, Virginia, United States: Logistics Management Institute.

MACHINERY CONDITION PROGNOSIS USING MULTIVARIATE ANALYSIS

Sheng Zhang, Melinda Hodkiewicz*, Lin Ma, and Joseph Mathew

CRC for Integrated Engineering Asset Management, School of Engineering Systems, Queensland University of Technology, Brisbane, QLD 4001, Australia

**School of Mechanical Engineering, University of Western University, Perth, WA 6009, Australia*

Abstract: Condition prognosis is an essential element for predicting the health of physical assets. Prognosis enables a projection of asset condition progression from the past and current into the future, thus providing a significant input into the asset related decision making process. Of different prognosis technologies, signal or feature based methods make predictive decisions according to progressive condition information and have been one of the focuses of recent research in the area. The success of these methods largely relies on effective feature extraction and feature manipulation. This paper proposes the use of eigenvector analysis for machinery condition trending and prognosis which incorporates multiple variables. An eigenvector analysis method, Principal Component Analysis (PCA), is selected to handle the features extracted from both time and frequency domains of vibration signals. The derived composite features facilitate the visualization of condition progression from different dimensional views, hence assisting condition prognosis. Control limits of the composite features are also designed for triggering condition alarms. The proposed method is verified using data from pump condition monitoring during flow fluctuations.

Key Words: Condition prognosis, Physical assets, Condition progression, Predictive decisions, Eigenvector, PCA, Pump, Flow.

1 INTRODUCTION

The research focus in condition monitoring has been shifting from fault diagnosis to prognosis in recent years. As fault diagnosis undertakes the investigation of past asset condition, prognosis targets the prediction of future asset condition. The ISO defines prognosis as "an estimation of time to failure and risk for one or more existing or future failure modes" [1]. Such an estimation derived by prognosis techniques is of significance for asset management. It assists in a range of asset related decision making, such as determining maintenance schedules, asset renewal, replacement, and disposal. Pursuing a capability in asset health prognosis has become an important agenda for many enterprises.

Conducting successful prognosis, however, is more difficult than conducting fault diagnosis. Prognosis generally requires a sound understanding of asset condition history. A much broader range of asset health related data, especially those related to the failures, shall be collected. The asset health progression can then be possibly extracted from the congregated data, which has proved to be very challengeable. Technically, prognosis can then be carried out by using predictive techniques to project this condition progression into future [2]. As a result of such challenges, prognosis is "normally intuitive and based on experiences", and relies highly on fault diagnosis, and "it is usually effective for faults and failure modes with known, age-related, or progressive deterioration characteristics, the simplest of which is linear. Prognostics are most difficult for random failure modes"[1].

A number of prognosis models have been developed in the past. These models can be roughly allocated into four categories, physical models, signal based models, reliability based models, and hybrid models [3][4]. The general ideas behind these models are somehow related - they more or less apply regression or extrapolation techniques to project history and current conditions into the future.

The literature contains several publications in the area particularly those that use signal based models for condition prognosis. Mba [5] conducted studies in the prognosis of gears and bearings. Pitting and spot defects were selected as the failure modes respectively. Vibration and Acoustic Emission (AE) were used as the basic technology for developing the prognosis algorithm. Wang etc. [6] studied the prognosis for gears with wear, chipped and crack failures using a Neuro-fuzzy method. Jantunen [7] studied the prognosis for bearings using accelerated experiments. The polynomial regression technique was applied to derive condition prediction. Byington and Roemer et al. [3][8] studied the gas turbine fuel nozzle. Clogging in the nozzle was selected as a failure mode.

Some work has been reported on physical model based prognosis techniques. Both Li [9] and Byington etc. [3] proposed a physical model for the simulation of crack growth in gears. Reliability based models are reasonably well advanced for prediction in maintenance. For prognosis, some efforts have been expended in the manipulation of condition data and reliability information to derive more accurate asset health prediction. Two models have been reported. One is the Proportional Hazard Model (PHM) which incorporates asset environment variables into the reliability model. This model has been comprehensively studied by Jardine, Vlok etc [10][11][12]. The Proportional Covariates Model (PCM) which combines response variables of asset condition into the reliability model is an additional approach. This model was proposed by Sun et al [13]. Apart from these models, Hess et al [4] suggested using hybrid models which combine different prognosis methods to reach a potentially more effective solution for prognosis.

Most of the works on prognosis have been conducted at a component level. Performing prognosis at a system level is relatively more complex. Some traditional methods, such as Reliability Block Diagrams (RBD) and Fault Tree Analysis (FTA), enable reliability analysis at the system levels and may be expanded for prognosis. Also, hybrid models can be investigated for system level reliability analysis and can eventually achieve a prognosis estimation.

This paper presents a signal based model for machinery condition prognosis using vibration signals from a pump during a flow fluctuation process. The vibration signal contains rich information from both time and frequency domains. This provides one with the opportunity for feature extraction using multivariate analysis. An eigenvector analysis method, Principal Component Analysis (PCA), has been used in the past [14][15]. It will be applied to trace the flow fluctuation process indicated by vibration features. This method is suitable to depict any condition progression regardless the failure mode.

The rest of the paper is organized as follows. In Section 2, the eigenvector analysis and PCA theory are briefly described. Section 3 presents multiple features extracted from vibration signals. In Section 4, the experiment on pump condition monitoring is described and vibration data is used to validate the proposed method. The results of the analysis are also presented in this section. The conclusion is drawn in Section 5.

2 EIGENVECTOR ANALYSIS AND PCA

Among all multivariate analysis methods, eigenvector analysis has attracted substantial interest. Of which the PCA is widely accepted for feature selection or feature compression. PCA transforms data from a variable space to a principal component space. The principal components containing large amount of data variability can be selected for condition monitoring purpose.

PCA approximates a general matrix by a sum of a few matrices ranking one. Let A be a real observation matrix of size $m \times n$ (n variables and each variable with m samples), then the economic-sized singular value decomposition on A is [16]

$$A = U \Sigma V^T \tag{1}$$

A can be rewritten as the sum of orthogonal component matrices

$$A = \sum_{i=1}^{p} E_i \text{, with } p = \min(m,n) \text{, } E_i = \sigma_i u_i v_i^T \tag{2}$$

where, σ_i are the singular values of E_i, μ_i and v_i are nonzero singular vectors from the ith column of U, and the transpose of the ith column of V, respectively. The norm of each component matrix E_i is its corresponding singular value σ_i. Therefore the contribution that E_i makes to reproduce A is determined by the size of σ_i. A can then be approximated by E which is the sum of r component matrices.

$$E = \sum_{i=1}^{r} E_i \tag{3}$$

The resulting E is called the loading matrix which has the size of $r \times p$ and rank of r. The importance of ith component, I_i, can be determined by comparing the ratio of its singular value and the sum of singular values.

$$I_i = \frac{\sigma_i}{\sum_{i=1}^{r} \sigma_i} \tag{4}$$

The correlation matrix is commonly used for A in principal component analysis. For example, given an observation matrix being standardized by subtracting the means of the columns and dividing by their standard deviations, A is the cross-product of the standardized observation matrix.

The transformation matrix E, which is derived from the observation matrix A, describes the mapping between the variable space and the principal component space. Hence, given X as a signal matrix with size of $p \times n$ (each row as a variable, and each column as a sample), the transformed matrix Z is

$$Z = EX \tag{5}$$

where, Z is of size $r \times n$, with each row a Principal Component (PC), and each column a sample.

The resulting principal components can then be selected as condition indicators. The top principal components are commonly selected since they contain a large portion of variations in the raw variables. However, there are no firm guidelines about how many principal components needs to be retained in order to produce an effective monitoring procedure. Since the principal components are the linear transformation of raw variables, they do not express physical explanation directly. This renders a problem for setting up alarm limits for the components. The alarm limit can be generally derived from the PCA analysis of accumulated observation data. The major components can also be plotted vs. each other. The resultant plot or trajectory indicates how the condition is progressive in the principal component space. The limit for the trajectory can be designed using multivariable control chart methods, where the principal components are fitted by a multivariate empirical reference distribution to establish a control region for the condition process [17].

3 MULTIPLE FEATURES FOR CONDITION MONITORING

Various data are generally available in condition monitoring, such as vibration, oil debris, thermograph, temperature, and pressure. Vibration is probably the most widely used signal in machine health assessment. Multivariate analysis e.g. PCA can be applied to the condition monitoring data or the vibration signal only for feature extraction and health assessment.

Observed vibration signals are the result of excitation from relevant machine components. Different features can be extracted from vibration signals to identify the machine health condition. These features are usually derived from either the time or frequency domains. Joint time-frequency domain signals, such as those developed using wavelet techniques, can be considered as time waveforms confined in specific frequency bands. A list of vibration features has been proposed and studied in [18], which includes P2P, RMS, Kurtosis, Crest factor, <1RPM, 1RPM, 2RPM, sRPM, >10RPM, OA, Band1, Band2, Band3, Band4, Band5 and Band6.

The first four features are vibration measurements from the time domain. They are largely the result of time waveform moments and their derivations by using statistical analysis. RMS, the first order of signal moments, reflects the signal energy and is an important indictor of vibration severity. Kurtosis, the fourth order of signal moments, reflects the peak characteristic. Crest factor is defined as the ratio between the peak value and RMS. Both kurtosis and crest factor are effective in detecting periodic impacts caused by the faults such as, rolling element bearing wear, gear tooth wear and pump cavitations, for which the spectrum analysis may fail to differentiate in the presence of random noise.

The rest features in the list are from the frequency domain. The feature frequencies, i.e. 1RPM, 2RPM and sRPM (the sum of harmonic and half harmonic between [0-10] RPM), are rotational speed dependent and can indicate the symptom of a specific component or fault. OA (Overall Spectrum) is an overall measurement of the signal in the frequency domain. The six band features are the measurement of signal components in specific frequency ranges. A reference definition of the band features is given by the Technical Associates of Charlotte [19]. Each of the six bands contains specific information sources. Band4, for instance, roughly covers the fundamental bearing defect frequency. While Band5 and Band6 cover the low and high harmonics of bearing frequencies, respectively. Note that the characteristic frequencies of specific parts, such as rolling element bearings, gearboxes, vanes and impellers, are also important condition indicators. For example, the characteristic frequencies of rolling element bearings are effective in identifying failures cased by faulty inner race, outer race, ball or cage. These features are generally machine structure and rotating speed dependent. One may care to add more features from time and frequency domains, such as the Spectrum Peak Ratio (SPR) [20].

Reference severity charts for monitoring the features have been proposed by the ISO and other practitioners. For example, the ISO 10816-3 (original the ISO 2374) gives a measurement of overall velocity in terms of RMS. Charlotte provides a chart for measuring overall velocity in terms of Peak to Peak (P2P) value and also suggests reference limits for the band spectrums. If no reference is available; the feature limit can be statistically derived from baseline data – a practical approach in most applications. A rule of thumb to determine the severity limit is double of the baseline for most vibration features.

The vibration features can be utilized individually in both diagnosis and prognosis. However, they are more often utilized in an integrated manner. There are two approaches to manipulate multiple features. One approach employs individual features for diagnosis or prognosis first. The conclusions from each individual feature are then fused by a decision fusion model to reach a final conclusion. This approach has been used in [21] for diagnosis. Another approach is to manipulate the features as a whole. The most relevant features or their derivations can be determined by a suitable model. In this approach, the feature and decision fusions are reflected in the chosen model. This approach has been investigated using the above listed vibration features for rolling element bearing diagnosis [22], where Bayesian neural networks were applied.

This paper investigates feature manipulation for prognosis. The application of PCA for feature extraction in this work falls under the second approach, which conducts the feature fusion in the PCA model.

4 EXPERIMENTS AND RESULT

Double-suction pumps are a special class of centrifugal pumps used in the process and water industries for high volume, medium head applications. Process plant experience has shown that partial flow operation can result in severe axial vibration and subsequent component failure. Therefore, it is valuable to gain an insight into the effect of partial flow on the axial vibration response of double-suction pumps.

This paper adopts the results from [23] to demonstrate the PCA application in prognosis. Vibration signals were collected from a double-suction pump (pump set A) shown in Fig. 1. The pump has two casing suction passages joined to a common suction nozzle and flange. The flow enters to the pump and is split into two annular suction chambers that turn and diffuse before entering the impeller through the opposing entrances. The flow fields from the two suction eyes are joined mid-way through the impeller vane passage and discharge into a common volute. The impeller is mounted on a shaft which passes through two suction chambers and is supported by two bearings. The bearings support the hydraulic loads imposed on the impeller, the mass of the impeller and shaft, and the loads from the drive components. They keep the lateral deflection of the shaft within acceptable limits for the impeller and shaft seal. The detailed specifications of the pump are given in [23].

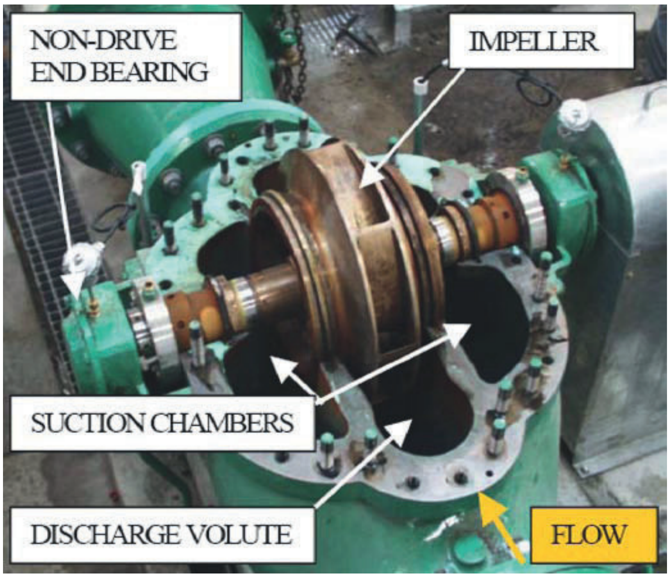


Fig. 1 The double-suction pump

When the impeller is centred in the casing and operated under design conditions, axial vibration levels are very low. Peaks at vane pass frequency are detectable anywhere on the pump casing and bearing surface and this is the major contributor to the pump's overall vibration. Radial vibration data measured on the bearing housing(s) in the horizontal and vertical directions contains vane pass frequency and often harmonic rotational components. During partial flow operation a complex, unsteady, asymmetric flow pattern at each suction eye results in loss of the hydraulic balance and axial motion of the impeller which can be detected as an axial vibration on the bearing housing of the pump.

The axial vibration of the pump was measured at 15 discrete operating points over a range of flows [0.40, 1.20] *Q*/*QBEP* (ratio relating measured flow rate to the flow rate at best efficiency point) by adjusting the position of a valve on the discharge line of the pump. The 15 operation points are listed in Table 1.

Table 1. Operation points of the pump

Test ID	1	2	3	4	5	6	7	8	9	10	11	12	13
Q/*QBEP*	0.42	0.56	0.59	0.67	0.8	0.84	0.99	1	1.01	1.11	1.14	1.2	1.2

At each operating point, 150 seconds of vibration data were digitally tape recorded using B&K 4396 Deltatron accelerometers stud mounted on the non-drive end bearing in the axial location. The data was sampled at a 2k*Hz* sampling rate when the pump was running at 1459RPM. Meanwhile, the performance data providing flow, pressure and pump efficiency was measured using a Yatesmeter [24].

The 15 operation points describe a progressive condition during flow fluctuation. The vibration signals of the condition can be processed by PCA and future vibration due to flow changes can then be predicted. The sixteen features as listed before were used individually to describe the vibration changing along the flow fluctuation. Fig. 2 demonstrates the trend depicted by each feature. As can be seen, some features show a clear trend e.g. P2P and RMS, while others do not e.g. Kurtosis and Crest Factor. There is a need to extract composite features to represent the trend from these multiple features.

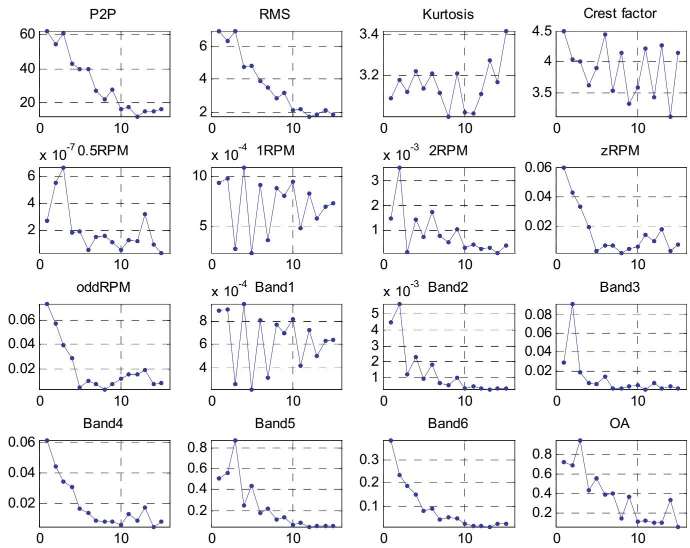


Fig. 2 Individual feature trending

The PCA model is applied to multiple feature manipulation. In theory, 16 principal components can be drawn from the feature sets. Fig. 3 illustrates the top four principal components – they represent the top ranking components which contain a large portion of variations in the raw variables. The result shows that the first principal component is very dominant compared with the rest since it contains 93% of the data variation (from Eq.(4)). Generally, the approach would be successful if sufficient components are included as a reasonable proportion of the total process variation. However, as mentioned before, there is no rule on how many PCs should be included. The components can be shown jointly to provide a process trajectory. Fig. 4 and Fig. 5 give the results in a 2-D and 3-D view respectively.

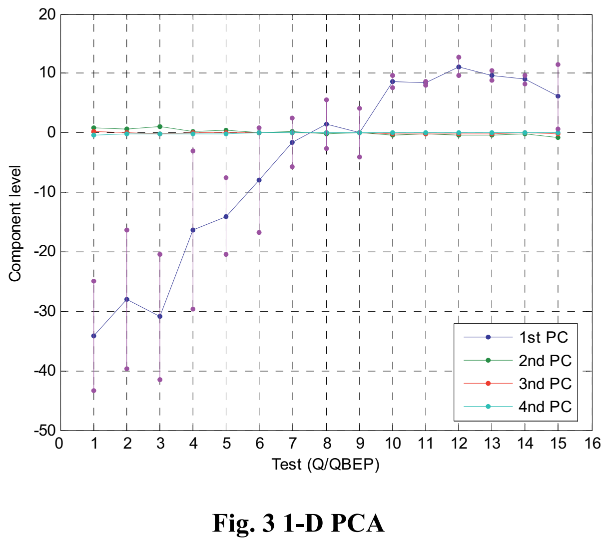


Fig. 3 1-D PCA

The three PCA diagrams present a clear trend of vibration changes during the flow fluctuation. This trend gives the opportunity for prognosis to predict future vibration condition due to flow fluctuation. Also, if the operation points near the Best Efficiency Point (BEP) of the pump are taken as normal conditions, e.g. operation points 7, 8 and 9, then the alarm limit can be plotted by a specified percent of confidence. In this study, the 2-D and 3-D principal components were assumed to follow a joint multivariate Gaussian distribution respectively. A 95% confidence for the normal condition consisting of the points 7, 8 and 9 was designed. This confidence level corresponds approximate 2.5 times of standard deviation of the multivariate in the Gaussian distribution [17]. Abnormal operation points can then be displayed which are beyond the normal confidence range as illustrated in Fig. 4 and 5. The change of vibration condition due to flow variation can then be traced easily from the diagrams.

In all the three diagrams, each operation point can also be described using a confidence boundary. Again, a 95% confidence level was used to describe the operation point in all the three cases.

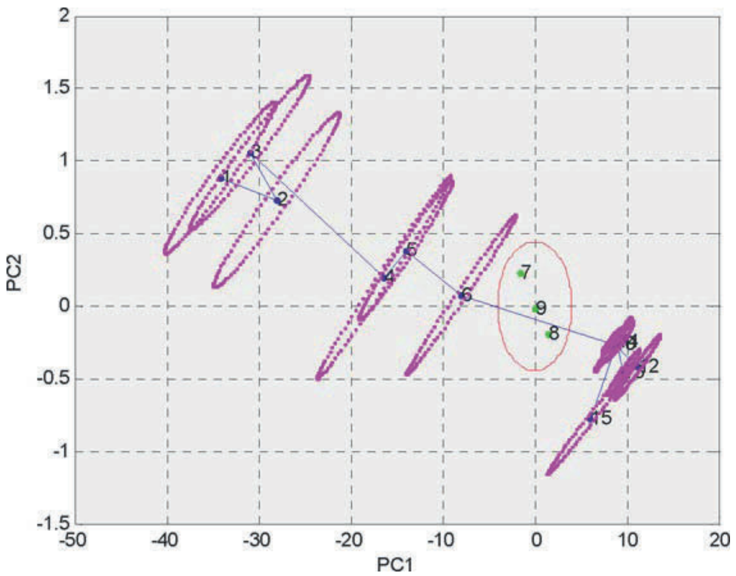


Fig. 4 2-D PCA

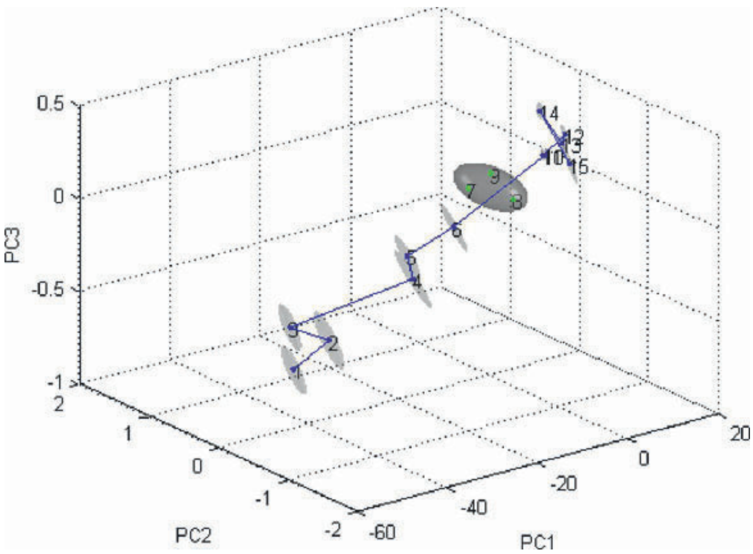


Fig. 5 3-D PCA

5 CONCLUSION

A multivariate analysis method, principal component analysis, was investigated in this work for condition prognosis purpose. It was applied to extract composite indicators (principal components) from different vibration features which are generally used for indication of condition variations. The different dimensional views, 1-D, 2-D and 3-D of principle components, facilitate the visualization of condition progression, thus ease the tracing of condition changes and assist prognosis. The vibration response from a double-suction pump during flow fluctuation was selected for the verification of the method in prognosis applications.

6 REFERENCES

[1] ISO 13381-1: Condition monitoring and diagnostics of machines – Prognostics.

[2] Jardine A. K. S., Lin D. M., Banjevic D., A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mechanical systems and signal processing, in press, 2006.

[3] Byington C. S., Roemer M. J., Galie T., Prognostic enhancements to diagnostic systems for improved condition-based maintenance, Aerospace conference proceedings, IEEE, 6, 2815-2824, 2002.

[4] Hess A., Calvello M. G., Frith P., Challenges, issues, and lessons learned chasing the “Big P”: real predictive prognostics Part 1, IEEEAC paper # 1595, 2005.

[5] Mba D., Prognostic opportunities offered by acoustic emission for monitoring bearings and gearboxes. Twelfth international congress on sound and vibration, 2005.

[6] Wang W. Q., Golnaraghi M. F., Ismail F., Prognosis of machine health condition using neuro-fuzzy systems. Mechanical systems and signal process, 18, 813-831, 2004.

[7] Jantunen E. Prognosis of rolling bearing failure based on regression analysis and fuzzy logic, Proceedings of the VETOMAC-3 and ACSIM-2004, 390-399, New Delhi, India, 2004.

[8] Roemer M. J, Kacprzynski G. J., Advanced diagnostics and prognostics for gas turbine engine risk assessment, IGTI/ASME Turbo Expo, May 2005, Munich, Germany.

[9] Li C. J., Lee Hyungdae, Gear fatigue crack prognosis using embedded model, gear dynamic models and fracture mechanics. Mechanical systems and signal processing, 19(4), 836-846, 2005.

[10] Jardine A. K. S., Banjevic D., Wiseman M., Buck S and Joseph T, Optimizing a mine haul truck wheel motors’ condition monitoring program: use of proportional hazards modelling. Journal of quality in maintenance engineering 7, 286-301, 2001.

[11] Vlok P. J., Coetzee J. L., Banjevic D., Jardine A.K.S. and Makis V., Optimal component replacement decisions using vibration monitoring and the PHM. Journal of the operational research society 53, 193-202, 2002.

[12] Vlok P. J., Wnek M., Zygmunt M., Utilising statistical residual life estimates of bearings to quantify the influence of preventive maintenance actions. Mechanical systems and signal processing, 18, 833-847, 2004.

[13] Sun, Y.; Ma, L., Mathew, J., Wang, W.Y., and Zhang, S., Mechanical systems hazard estimation using condition monitoring, Mechanical systems and signal processing, 2004,12, available on-line

[14] Cempel C., Multidimensional condition monitoring of mechanical systems in operation, Mechanical systems and signal processing, 17(6), 2003, 1291-1303.

[15] Cho M. S., Kim K. J., Indirect input identification in multi-source environments by principal component analysis. 16(5), 873-883, 2002.

[16] Stewart G. W., Matrix algorithms: basic decompositions, SIAM, 1998.

[17] Montgomery D. C., Introduction to statistical quality control, John Wiley & Sons, New York, 2001.

[18] Zhang S., Ma L. and Mathew J., Vibration feature analysis for rolling element bearing diagnosis using Bayesian neural networks, 11th Asia Pacific vibration conference (APVC), Langkawi, Malaysia, 373-378, 2005.

[19] Website of Technical Associates of Charlotte, P.C. - http://www.technicalassociates.net/.

[20] Altmann J. and Mathew J., Multiple band-pass autoregressive demodulation for rolling element bearing fault diagnosis. Mechanical systems and signal processing 15(5), 963-977, 2001.

[21] Zhang S., Mathew J., Ma L. and Sun Y., Best basis based intelligent machine fault diagnosis, Mechanical systems and signal processing, 19(2), 357-370, 2005.

[22] Zhang S., Mathew J., Ma L. and Sun Y., Identifying optimal combination of wavelet packet bases for fault classification using Bayesian neural networks, Proceeding of VETOMAC-3 AND ACSIM-2004, 1231-1237, New Delhi, India, 2004.

[23] Hodkiewicz M. R. and Pan J., Identification of transient axial vibration on double-suction pumps during partial flow operation. Acoustics Australia, 32(1) 25-32, 2003.

[24] Hodkiewicz, M.R. and Norton M. P., The effect of change in flow rate on the vibration of double suction centrifugal pumps. Proceedings of the institution of mechanical engineers Part E: process mechanical engineering, 216, 47-58, 2002.

STRATEGIC ASSET MANAGEMENT
– GETTING THE PEOPLE AND SYSTEM ISSUES RIGHT

Albicini, V[a], Marmaras, F[b] and Hanrahan, M[a]

[a] *Knox City Council, Burwood Hwy, Wantirna South, Victoria 3152, Australia.*

[b] *Integr8 Pty Ltd, Suite 303,145 Russell Street, Melbourne, Victoria 3000 , Australia.*

Abstract: This paper details the recent experience of a Victorian Local Government authority, Knox City Council, as it attempts to implement strategic asset management principles across the organisation. It outlines challenges faced when shifting from a culture that supports a reactive response system to one that encourages a proactive, whole of network approach that uses risk to prioritise workloads. Knox encountered two main challenges in delivering on Council's strategic asset management objectives – getting the people issues right and getting the system issues right. Whilst Council recognises that it still faces considerable challenges on both fronts, it has come to realise that people and system issues are interdependent and cannot be analysed or improved in isolation.

Key Words: Strategic Asset Management, Change Management, Asset Management Systems

1 OPERATING ENVIRONMENT

Knox City Council is an outer eastern Municipal Council in Melbourne. The Operations Centre at Knox Council is responsible for completion of maintenance and renewal works on engineering infrastructure across four work areas – Parks (active and passive reserves, vegetation, playgrounds), Works (roads, paths, drainage), Buildings (community facilities, child care centres, pavilions) and Fleet (motor vehicles, plant). This paper focuses on the changes necessary to improve operational output in both the Works and Parks areas through application of strategic asset management principles. To achieve performance improvements, it was necessary that changed work practices be implemented at a procedural level (the system used to record and manage work activities) as well as at a people level.

Figure 1. City of Knox

Traditionally, Operations Centre staff have operated on a reactive basis, with performance measurement systems that encouraged a response to the squeakiest wheel rather than an objective assessment of need. A customer request system which channels requests from the community or Council officers to the Operations Centre was the key driver of maintenance action within the municipality. Timeframes for asset repair typically reflected resource and funding availability, with little documentation available to quantify works undertaken, plant and labour requirements and justification for carrying out works. Typically, the decision to undertake maintenance works was solely at the discretion of the field officer, administrator or their supervisor. Review, auditing and quality assurance were undertaken but not to any great degree.

Table 1 contrasts the people and systems at Knox today with the situation before the journey toward strategic maintenance service delivery was initiated.

One of the more significant challenges involved changing the people culture within Knox, with Operations Centre staff often seeing themselves solely as a maintenance crew, finding little correlation between their own works and the performance of the asset as a whole. This was in part a result of deficiencies in the asset management system and appropriate reporting mechanisms that could reliably inform staff about asset performance. This itself was likely a bi-product of the 'as needed' relationship which had developed between Council's Engineering, Asset Management and Operations Centre departments, providing minimal opportunity for effective interaction between work units.

Table 1: People and systems before the journey commenced and now

PEOPLE - Maintenance crews and supervisors	
Before the Change	**Now**
– administrators drove the service instructing field staff to address customer requests – isolated from other departments responsible for asset management (planning, design, construction, renewal, disposal) – inconsistent responses to customer requests – delivery times for repair not objective – status quo not challenged – officers saw themselves as doers not thinkers – reluctant to become involved in strategy development – confident in ability to repair assets but lacking skills in data recording and analysis	– recognise and communicate to others the benefits of a structured proactive approach to asset management – supervisors drive the service instructing field staff on better ways of delivering the service – consistent responses to customer requests based on a robust risk assessment process to justify action or in-action – earning respect from other departments because they are able to quantify and justify what they do – more confident to challenge traditional work practices – eager to support expansion of system into other interrelated departments – confident in ability to repair assets improving data recording and analytical skills
SYSTEMS - Computer applications and work prioritisation processes	
Before the Change	**Now**
– asset management system did not record any maintenance data against assets – independent customer response and finance systems – no maintenance management system – no documented maintenance intervention levels to dictate whether an asset requires repair – customer response system used to initiate maintenance activities. Decision to repair asset at discretion of crews, administrators and supervisors – limited inconsistent recording of maintenance actions undertaken by field staff – adhoc asset inspections used to supplement workloads during quiet times – no clear auditable process for prioritisation of workloads - multiple spreadsheets managed by multiple supervisors with inconsistent, overlapping, incomplete information – no centralised codified system – difficult to locate and interrogate data – performance auditing difficult due to lack of documented work processes and responsibilities and data collection in disparate systems	– maintenance data recorded against individual assets enabling sites with repeat issues to be easily identified and future prediction of asset deterioration rates – new maintenance management system integrated with existing asset management system, customer response and finance systems – documented maintenance intervention levels dictate whether an asset requires repair and timeframe for completion of repair based on risk – maintenance service levels built into the new maintenance management system to drive changes in work practices and measure performance – issues raised by customers assessed to determine whether they exceed intervention levels. Timing for completion of repair, if required, based on public safety risk assessment at the time of initial inspection – regular asset hazard inspection program introduced. Assets with defects that exceed maintenance intervention levels recorded, public safety risk assessed at each location. Work prioritised and delivered accordingly – adhoc asset inspections enable crews to report high public safety risk issues identified on site – centralised codified system - data can be analysed to determine workloads, causes of asset failure, performance in delivery of target repair timeframes , reasons for delays etc. – auditable work processes based on public safety risk

At a senior management level, strategic asset management had been identified as a core function of Council's future operations, however responsibility for the development of strategic plans at this early stage rested solely with a newly created Asset Management department. Implementing the strategic objectives and the required cultural and procedural changes in a manner which was meaningful and applicable to staff rested with the Operations Centre Manager with support from the Asset Management team. It was recognised that a system alone would not be able to deliver the desired results, but also noted that a change process without a key driver (system) would be extremely difficult. The considerable task of system development coupled with people management was subsequently initiated.

It is important to note that enacting change in staff practices within a public sector environment has its own challenges. Organisational inertia is compounded by public and political scrutiny and the financial and operational constraints that these may bring. Major changes to the structure of the organisation and staff position descriptions are difficult. Skills and capabilities of staff may not be completely understood or utilised effectively in an environment where monitoring and reporting has not previously been well utilised. Effective senior management is therefore essential to not only develop the vision for the team, but also be prepared to take all steps necessary to deliver this vision in the face of any barriers which may present themselves.

2 DRIVERS FOR CHANGE

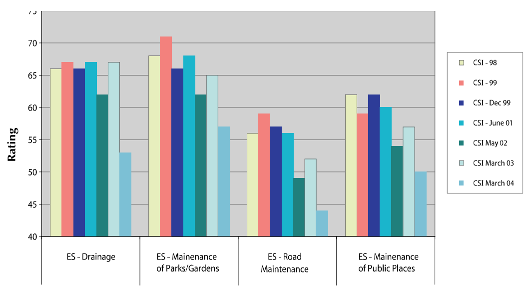
2.1 Customer Dissatisfaction

The original Knox shire extended over 110 square kilometres and included the suburbs of Bayswater, Boronia, Ferntree Gully, Knoxfield, Wantirna, Wantirna South, Rowville, Scoresby and The Basin. At the time of its inauguration on 16 November, 1963, the Shire of Knox had a population of just 24,000. During the 1970's and 1980's the population density of the City grew with an influx of businesses and rapid expansion in the residential sector. Today 147,000 people live in Knox.

Figure 2. Council Assets

Roads	***703 km***
Kerb & Channel	**1 300 km**
Footpaths	**1 166 km**
Bike Paths	**70 km**
Drainage	**1 107 km**
Playgrounds	**203**
Buildings	**262**
Parks/ Gardens	**650 ha**

By the early 90's as growth of the municipality had begun to slow it was becoming obvious to Council's Senior Management team that Council faced a huge potential future liability if its assets were not managed for the long term. Council was left responsible for maintaining an ageing stock of infrastructure assets to a standard deemed acceptable by its customers – the same customers who possess ever-increasing expectations for improved services and an ever-reducing willingness to pay. Surveys undertaken by Council of residents and businesses indicated that the community was often unhappy with the way key assets such as footpaths, roads, drains and parks were being managed.



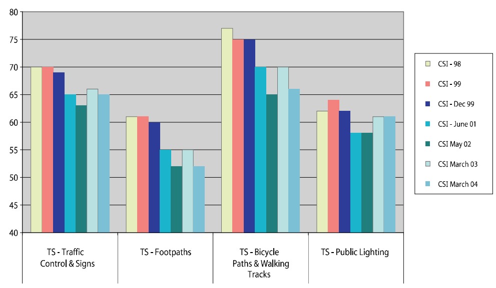


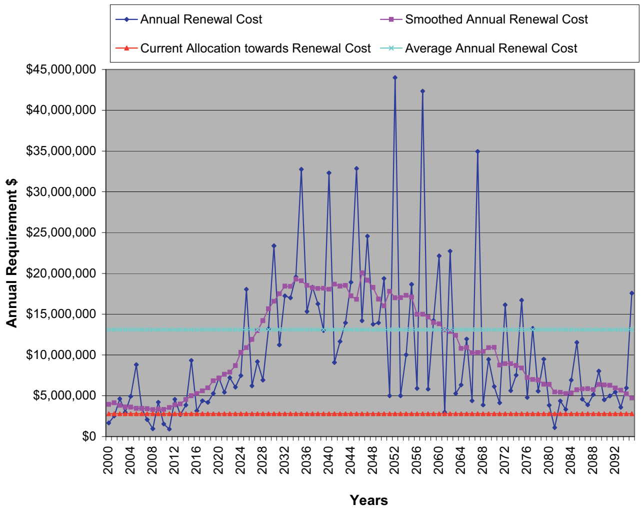
Figure 3. Auspoll – Customer Satisfaction Survey Results

2.2 Staff Recognition of Inefficiencies in Service Delivery

Internally, there were also some concerns, with staff satisfaction surveys suggesting a need to review internal processes to improve communication among departments, remove duplication of effort in some areas, and provide greater opportunity for professional development. In addition to this, the Operations Centre underwent a major review of its service delivery areas in mid 2004. This review assessed requirements and capabilities of all Operations Centre staff, structure issues and service delivery improvement opportunities. The summary findings of this review were seen to be a key driver for the change needed at Knox, with staff participation in this review raising awareness of key issues and setting the environment for future change.

2.3 Financial Sustainability

Council initially responded to some of the above concerns with the development of a Strategic Asset Management Plan. This strategic document researched and assessed Council's assets in an attempt to broadly quantify the age, condition and performance of Council's asset stock and subsequently predict the extent of future liabilities under varying funding scenarios and demand/growth projections. The data was to provide a basis for informed decision making which would assist with long term financial planning in Council. The Plan took into account the need for clear decision making processes associated with service provision, risk management, future asset growth and rationalisation opportunities. The framework for delivery of the plan was linked directly with Council's commitment to sound asset management, which had been articulated in Council's long term vision and Corporate and Community Plan.



Asset Category	Current Replacement Cost ($,000)	Valuation (30/6/05) ($,000)
Footpaths	87,857	41,059
Shared-use Paths	8,203	3,180
Drainage	184,896	124,863
Kerb & Channel	93,199	58,512
Road Surface	49,000	14,290
Road subsurface	183,290	156,071
Roads – Earthworks	119,204	119,204
Bridges	4,740	3,548
Buildings	116,460	76,070
Playgrounds, park equipment & furnishings	22,470	14,726
Bus Shelters	167	129
Carparks	10,467	7,038
Land	292,492	292,492
Plant	6,846	5,102
Office Furniture & Equipment	5,638	1,466
Artworks	223	223
InProgress (30/6/05)	584	584
Total	1,191,007	923,329

Source: 2004/05 Annual Report

Figure 4. Financial liability – ageing assets following a peak growth period

This strategic document was used to communicate and influence elected officials on the importance of embedding strategic asset management principles into day to day decision making. However, like many strategies, staff at Knox were best placed to apply these principles through their own work practices. The magnitude of funding required to bring all assets to an acceptable service level in the eyes of the community using only limited funds would require risk-based prioritisation of maintenance and asset rehabilitation expenditure. In order to justify and enforce such decision making, improved systems would be necessary to adequately record and quantify the performance of Council's assets.

2.4 Legislative Change – The Road Management Act, 2004

The momentum for change did not gather pace until a sense of urgency was created by the High Court decision (Brodie vs Singleton Shire Council and Ghantous vs Hawksbury City Council) which made it possible for a road authority, such as Knox Council, to be found negligent for failure to act to inspect or repair a road asset. This decision ultimately resulted in the development of the Road Management Act, 2004 in Victoria, legislation which provided senior management with a motive and opportunity to drive change in the maintenance delivery methods of Council.

The legislated deadline for compliance being a period of only 12 months meant that at a minimum, Council's asset inspection regime and maintenance service levels had to be developed and documented with staff educated and geared to deliver and record the delivery of these. This tight timeframe for delivery of significant change meant there was little time for detailed forward planning. The team launched headlong into implementation of a more strategic approach to asset maintenance.

3 THE JOURNEY

The asset management team set themselves the task of delivering three key outcomes:

1. Develop and document maintenance service levels applicable to Council's drainage, park, road and road related assets;
2. Change service delivery principles from a reactive approach that serviced the most vociferous customers at the expense of others to an objective public safety risk based approach that considered each asset repair request within the context of the condition and criticality of the whole network of assets;

3. Upgrade of the existing asset management system to incorporate a maintenance management system that would support the changed service delivery principles, record and report on compliance with documented maintenance service levels and provide data for analysis to improve asset management into the future.

Delivering on the objectives required shifting the mindset of a team of some sixty maintenance staff, administrators and supervisors from a culture that supported a reactive response system to one that encourages a proactive, whole-of-network approach that uses risk to prioritise maintenance workloads. All tasks were to be completed within the surrounding legislative context with recognition of the need to maintain existing service to the Community at all times through the process. The key tasks undertaken are outlined in the table below

Table 2: The Journey – Key Tasks

Project Planning

- Education in principles of strategic asset management - limited to the Asset Management Team
- Informal planning of the change project aimed at overcoming hurdles as they were encountered.

Capture & challenge local knowledge to develop and document maintenance service levels

- Numerous formal meetings with officers responsible for asset repair to gather existing documentation, determine what they do and how work processes could be improved
- Maintenance service levels documented - including risk assessment to determine appropriate repair timeframes

Introduce changed maintenance service delivery principles

- Process developed for use by field staff to assess the risk associated with defects identified on site
- Field staff training :
  - assessing and recording public safety risk,
  - detailing the expected scope of work required to repair a hazard (including plant, labour and material requirements)
  - recording details of work undertaken
- Monthly meetings to discuss issues encountered by staff as they attempt to deliver documented service levels
- Performance auditing with measurable improvement requirements to drive change in work practices including data collection

Develop and introduce a new maintenance management system

- System developer engaged to create a maintenance management system (MMS) as an add-on to Council's asset management system.
- System development - not staged or tested
- System implementation –not staged, system required ongoing modification to iron out glitches, keep system functional and develop reporting tools that met Council's needs
- System training - informal "on the job" training for all users individually as system was rolled out.
- Monthly meetings with user groups to identify and prioritise system issues and identify improvements to be included in subsequent system upgrades

4 KEY LEARNINGS

In total, Knox have taken nearly two years on this journey toward proactive strategic asset management. Looking back, it is reasonable to surmise that the practicalities of delivering both system changes and people changes were significantly underestimated both in respect of time and effort to achieve the desired outcomes. Nonetheless, the journey provided us with valuable insights into staff management issues and the impact these may have from an operational perspective both now and into the future. In practice, senior management are now far better positioned to implement any future change within Council.

4.1 Espouse Strategic Asset Management principles to affected staff

Senior management created an asset management team and empowered them to challenge the status quo within the Operations Centre. This team were encouraged to build their knowledge by attendance at conferences and active participation in numerous asset management reference groups. This in itself created a divide between them and the Operations Centre staff, who had little initial awareness of the principles of asset management. As a result, initial discussions between the two teams left Operations Centre staff somewhat reluctant to reveal key data out of fear that such information may be used against them. Operation Centre staff had a great deal of pride in the way they had delivered the service on the ground and found the idea of being told by outsiders how to run their service objectionable.

A key challenge in the development of the maintenance management system was to ensure that staff saw any new application as being a permanent part of doing business and not simply a passing fad or strategic direction which had little application. General scoping of system requirements therefore involved Operations Centre staff. At this early stage however, the staff were not sufficiently educated in the principles of asset management to provide appropriate guidance for the consultant engaged to develop the system.

Early introduction of more staff to the principles of asset management by encouraging attendance at relevant conferences and participation in reference groups may have reduced the knowledge disparity between the two groups, provided the system developer with more coherent objectives, and created more widespread excitement toward the prospect of changing the way Council manages its assets. Even though many staff may perceive attendance at these sessions to be a waste of time better spent on productive activities they tend to provide an opportunity for staff to learn from others and challenge traditional ways of working in an unthreatening environment.

4.2 Plan and communicate the change management project

The twelve month timeframe for delivery of the three key project objectives meant that virtually no time was spent in the project planning stage. Planning including a stakeholder threat & opportunity analysis would have armed the team with a better understanding of the people issues likely to be encountered and provided an opportunity to develop systems and strategies to address them. In future, the team should have more private talks with as many people as possible, as early as possible to identify where support and challenges are likely to come from. Stakeholder analysis should identify whose cooperation is essential and key allies that can be relied upon to provide insight into others and later help drive the change effort by priming others to at least listen then follow their example. Allies don't have to be those in formal positions of authority but must be respected and trusted by other staff. The analysis of all affected individuals should consider the following:

- What motivates them, excites them and annoys them
- their concerns/ interests / issues
- how has change been avoided by them in the past
- expected type/ amount of resistance
- expected reasons for resistance
- their tolerance for ambiguity
- their likely allies (common goals)
- their likely enemies

This enables the team to group individuals with similar personalities, issues and tolerances to change together and then for each group, determine appropriate communication methods, specific tactics (structural, system, training, incentives etc) to counter predictable resistance.

It is important to note that affected individuals exist outside the affected department, and can be identified by looking at all work processes that involve interaction with other departments. Unfortunately, interdepartmental relationships were largely overlooked by the team. Customer service operators, charged with the responsibility of explaining likely maintenance repair works and timeframes for service delivery to customers found themselves providing customers with outdated information. Design engineers who had previously been able to direct the activities of maintenance crews, found themselves frustrated by crews that now operated according to a new set of guidelines based on a risk assessment to prioritise their workloads. This lack of communication was addressed by the team as a matter of urgency but could have been avoided by recognising and communicating changes to these indirectly affected parties. The above planning process would have revealed the importance of communicating maintenance service level changes.

4.3 Build trust to capture and challenge local knowledge

Maintenance service levels were established and now act as service delivery targets. They define performance expectations and were formulated based on local knowledge through an assessment of legislative requirements, organisational objectives, customer expectations and financial constraints. Staff at the Knox Operations Centre had a proud history of delivering continued service when delivering maintenance outcomes. Indeed, many staff at supervisory level had in excess of twenty years experience at Knox - resulting in detailed knowledge of asset performance and local area issues. Unfortunately, most of this knowledge rested within the minds and memories of staff, who were able to articulate when and where work was completed but without documentation to prove this.

When originally asked to quantify how they undertook their works in an effort to establish preliminary service levels, staff felt threatened and challenged by senior management. This may have related to a sense that they were failing to deliver an appropriate service or perhaps a more real and greater fear that, in documenting service levels, individual officers would personally be more exposed should they fail to meet such delivery timeframes or provide indicators of proof that the works had been completed. Previously, officers had received the benefit of a customer response systems which would allow timelines to be extended should a deadline not be met, resulting in a remarkably effective service delivery outcome.

D-066 - Damaged Pit Lid/ Structure
For all Drainage Pits within Road Reserve and along Shared paths:
Intervention Level

D-066A.	Pit lids, grates, and /or pit surrounds that are damaged or deteriorated to the extent that they pose a potential hazard to road users/pedestrians
D-066B.	Vertical displacement >15mm only if the pit is within a designated pedestrian walkway
D-066C.	Cracks >5mm considered likely to cause the pit lid and/ or surround to collapse prior to the next routine hazard inspection
D-066D.	Broken or missing pit covers
D-066E.	Broken frames that no longer support the pit lid
D-066F.	Missing/damaged / deteriorated step irons and/or mesh panels.

ACTIVITY TITLE:	Drainage Pit Lid/ Structure Repair (excluding lintels)
DESCRIPTION:	Repair or replacement of pit lids & structural components i.e. pit lids, grates, frames, step irons, surrounds, & mesh panels (installed to divert litter). (This activity includes drainage pits within the road reserve and along the footpaths/ shared path network).
POTENTIAL RISKS IF ACTIVITY IS NOT UNDERTAKEN:	Personal injury from falling into pits with missing lids/ grates/unstable frames Personal injury/ vehicle damage resulting from pit surface components not flush with road/ path surface Personal injury to workers due to broken/ missing step irons

REACTIVE MAINTENANCE	D-REA-066 - Drainage Pit Lid/Structure Repair (excluding lintels)
CURRENT SERVICE LEVEL:	Provide temporary and/or permanent repair when: a) Pit covers are broken or missing b) Cracks >5mm are considered likely to cause the pit lid or surrounds to collapse c) Vertical displacements >15mm within designated pedestrian walkways only d) Pit lids, grates, and/or pit surrounds are damaged or deteriorated posing a potential hazard to road users/pedestrians.
TARGET TIME FOR INITIAL RESPONSE: (Site Inspection/Public Safety Risk Assessment)	3 days
TARGET TIME FOR RECTIFICATION WORKS:	96 days

Figure 5 Sample maintenance service level

Given the short timeframe available to determine and document maintenance service levels there was little time available for the team to develop a trusting relationship with the knowledge holders before bombarding them with questions on what they do, why, how long it takes, and what could be done to become more efficient. Aware of this, the team made a conscious effort to build trust by acting in a consistent manner in all interactions with the knowledge holders. Delivering on responses to issues raised within agreed timeframes, not claiming to have all the answers, being forthright about their own mistakes and those of others, admitting their own feelings of uncertainty and asking directly about the feelings of others.

One of the most important recognitions for the team when developing service levels were that effective communication enabled greater clarity as to when maintenance work would be carried out by staff. Site visits, common language and sample photographs built trust and enabled adequate documentation of the service levels to be carried out. Immediate documentation of all thoughts and ideas was essential prior to the collation and review of the available data. Importantly, although a large quantity of material required review, documentation enabled all stakeholders to develop a common understanding and focus on reaching agreement on service levels to be delivered.

4.4 Assisting people to change – skill development and confidence building

The introduction of maintenance service levels that specified a timeframe for defects to be inspected, made safe and rectified based on a public safety risk assessment at the time of inspection meant that field staff needed to be able to understand the service levels and assess the risk that the defect would likely impose on public safety if no maintenance were undertaken. In addition field staff were for the first time required to record specific details of their assessment and the work undertaken so that this data could later be analysed to measure performance against service delivery targets and highlight potential process improvements.

The team recognised the fact that despite the legislative pressures for total compliance with documented service levels staff couldn't be expected to learn the required skills overnight. Intensive training, support and both positive and negative reinforcement were required. Those displaying clear resistance to change were actively exposed to the process and the logic behind the change and made accountable for delivering outcomes.

Knox City Council Works Order: 511287 *Date Received:* 25/01/2006 *CRS No:* 248850

Issue Location: 108 Rickards Avenue North KNOXFIELD 3180 *Extra Loc:* *Mel Ref:*
Customer Name Colin Watson *Ph No:* *GIS Ref:*
Customer Comments: BROKEN STORM WATER PIT LID/ 108 RICKARDS NTH AVE KNOXFIELD
Work Required: D-REA-066 Drainage Pit Lid/ Structure Repair (excluding lintels)
Service Level: Provide temporary and/or permanent repair when: a) Pit covers are broken or missing b) Cracks >5mm are considered likely to cause the pit lid or surrounds to collapse c) Vertical displacements >15mm within designated pedestrian walkways only d) Pit lids, grates, and/or pit surrounds are damaged or deteriorated posing a potential hazard to road users/ pedestrians.

☐ *Contact Customer Prior To Works* ☐ *Photos Req* ☐ *Insurance Claim Report Req* ☐ *Report Completed*

Temporary Works (Including Inspection): *Target Time for Initial Response:* 31/01/2006 1:52:35 PM
Allocated To: Greg Backhouse *Date Received:* 25/01/2006 1:52:35 PM
Contractor Fax No.: *** Contractors please quote Purchase Order Number on invoice ***

HAZARD:	CAUSE:	SAFETY RISK ASSESMENT:
☐ D066-A Pit Lids damaged to the extent that they are hazardous to road users/ pedestrians	☐ A001 - Accident	☐ *Extreme* ☐ *High* ☐ *Med* ☐ *Low*
☐ D066-B Grates damaged to the extent that they are hazardous to road users/ pedestrians	☐ A002 - Ageing Asset	**WORK REQUIRED:**
☐ D066-C Pit surrounds damaged to the extent that they are hazardous to road users/ pedestrians	☐ C001 - Inappropriate Construction	*Temp Protection Works* ☐
☐ D066-D Vertical displacement >15mm only if the pit is within a designated pedestrian walkway	☐ D001 - Inappropriate Design	*Rectification Works* ☐
☐ D066-E Cracks >5mm likely to cause the pit lid and/ or surround to collapse	☐ D002 - Drainage Inappropriate	**REFER TO:**
☐ D066-F Broken or missing pit covers	☐ D003 - Damage by Resident/ Builders	***Routine*** - N/A ☐
☐ D066-G Broken frames that no longer support the pit lid	☐ M001 - Inadequate Preventative Maintenance	**Renewal** - ☐
☐ D066-H Missing/ damaged/ deteriorated step irons and/or mesh panels.	☐ M002 - Maintenance Vehicle Damage	***Other Department*** -
☐ D066-Other	☐ S001 - Storms	Assets ☐
	☐ S002 - Subgrade Subsidence	Conservation ☐
	☐ T001 - Tree Roots	Operations ☐
	☐ U001 - Utilities	Engineering ☐
	☐ V001 - Vandalism	Local Laws ☐
	☐ V002 - Vehicle damage	Recreation Services ☐
	☐ Z001 - Other	Community Services ☐
		Other ☐

Comments *** (What Plant, Labour, Materials & OH&S Req. are needed to complete works.) Comments if referring works to Others. ***

Inspection Finalised By: *Signature:* ____ /__/__ __:__ AM / PM

Temporary Works Undertaken: *Target Time for Temp Works:* *Est Rect Cost:* $
☐ None ☐ Inspection Only ☐ Rectified Hazard
☐ Coldmix Fill ☐ Wedge ☐ Warning Signs ☐ Warning Lights ☐ Safety Fencing Installed
☐ Closed Asset to Public ☐ Other...
Temp Crew Members:
Temp Plant Used:
Temp Material Used: *** Include Quantities ***
Start Date & Time: __/__/__ __:__ AM / PM *Finish Date & Time:* __/__/__ __:__ AM / PM
Temp Works Finalised By: *Signature:*

Figure 6 Form used to assess issues

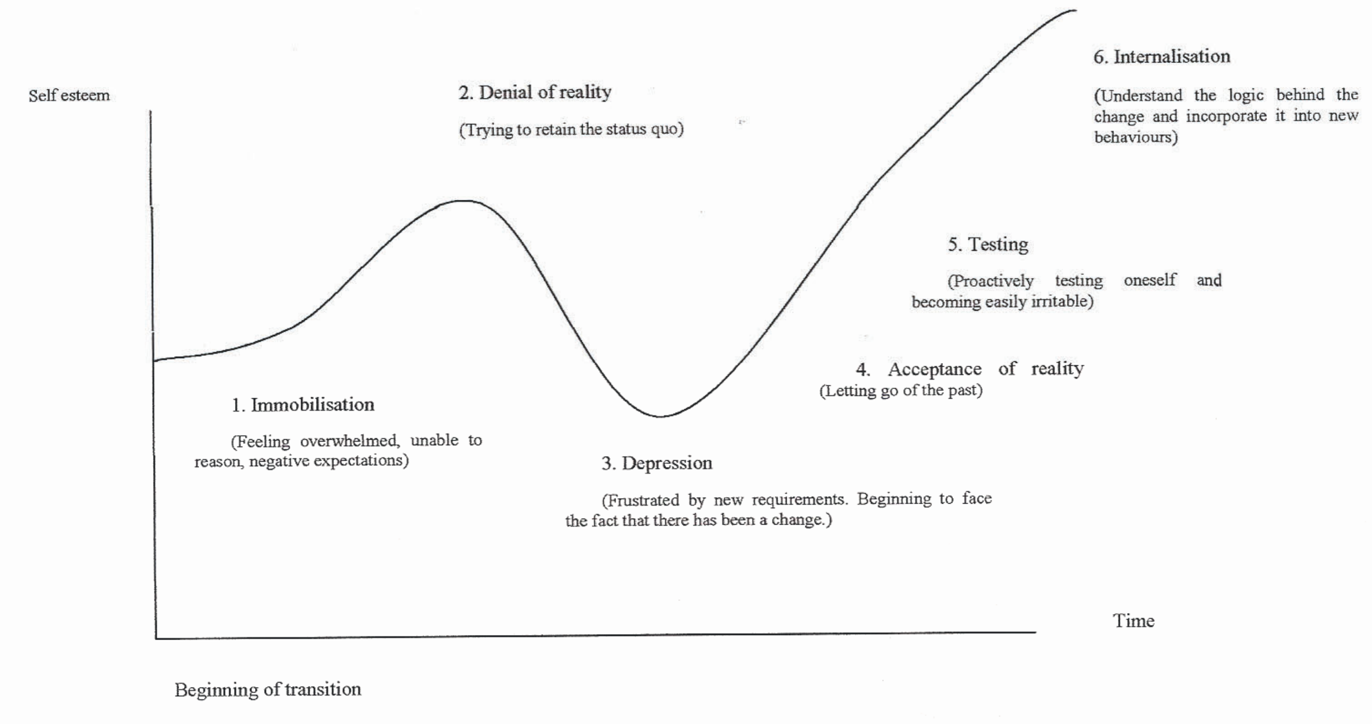


Figure 7. The emotional roller coaster of a changing work environment

A flowchart of the risk assessment process was documented as simplistically as possible and issued to crews for use on site. Supervisors were trained first, then assisted with workshops for field staff. Each training exercise was a hands-on application of the process rather than hypothetical or conceptual. The sessions involved real examples to give substance to the theory documented and provided to staff in both written form and flowcharts Training workshops acknowledged that the system may not be perfect and encouraged constructive criticism. These workshops provided information, skills and tools which were supplemented by a full time resource dedicated to working with groups and individuals to guide them in the new way of approaching issues as they tried to negotiate their day-to-day activities with a new set of rules. This support included open discussion to reinforce the fact that there was no hidden agenda, repetition of the change message to avoid staff slipping back to old familiar comfortable ways, allowing individuals to make mistakes and learn from them.

The collection and challenge of existing information and processes, together with the introduction of new work management and data collection processes and computer applications meant that at anyone time Council officers found themselves in one or more of the emotional states described in Figure 7. Successful change required the change management team to understand these emotional states and help officers move through the phases toward internalisation of the change.

Future change management should continue to include efforts to equip individuals with the skills to manage change. Informal coaching of staff would be enhanced by formal personal development programs aimed at equipping staff with an understanding of how and why people resist change, how to self evaluate, and develop coping mechanisms to deal with the emotional roller coaster that accompanies a changed work environment.

4.5 Plan system development and implementation

Given that delivery of the documented service levels could only be ensured if staff followed quite specific work processes it was recognized that a custom built system would be required. It was also considered essential that the maintenance management system be linked to Council's existing asset management system so that the data could be used in predictive lifecycle analyses. A custom built system was preferred because it could be designed to support the way we wished to operate. The key desired features included:

- systematic collection of maintenance histories against the affected asset
- ease of use and reporting
- integration with other corporate databases and applications (Latitude GIS and AXS One, GEAC Pathways)
- flexibility to change to reflect new reporting and business process needs
- ability to be expanded for use on portable handheld computer devices
- data validation to force users to use the system correctly

It was decided that the maintenance management system would be developed as an add-on to Council's existing asset management system by the original system developer with whom Council has a long relationship. Unfortunately it took longer than expected to develop a stable system relatively free of glitches. The delays were the result of a multitude of factors primarily:

- incomplete scoping of Council's needs at the outset
- work processes were evolving in parallel with system development as the team developed its understanding of how data should be collected so that compliance with service levels could be measured
- an unrealistic expected application delivery time
- system provider engaged to develop the system were moving their business focus away from provision of specialised system modules

Council's relationship with the system provider established over many years although tense enabled modification to the scope without a blow-out in costs.

Key system features are now inclusive of data validation capabilities, a system for recording worksite traffic management, a module for scheduling rectification tasks based on location, due date, risk and work activity, multiple reports including causes of asset failure and locations receiving duplicate requests. Performance reports are also able to be prepared to measure maintenance crews' timeframes for initial response to customer requests, temporary works to make the sites safe and rectification works and to measure the volume of work generated. Operations Centre staff and the asset management team met regularly during system development and continue to meet every three weeks to see how things are going. Issues continue to come up and are either resolved, a work around provided, or the need for a major system rethink noted for consideration when the system is next upgraded.

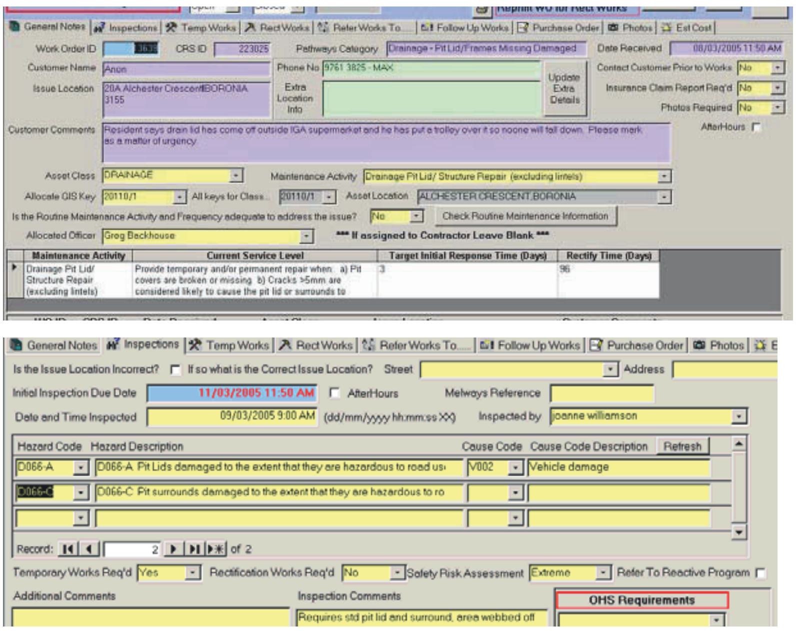


Figure 8 Sample data collection forms

Tight time constraints imposed by legislation, meant the team had no time to dedicate to the search of a system developer and little time to prepare a detailed scope as work processes were continually evolving. In hindsight however, as most people need little encouragement to avoid the confronting the challenge of actually trying to change the way they work the reliability of the system was critical to success and its development and implementation warranted a more thorough approach. Staff feedback on their experience of the change process, twelve months after the system was rolled-out has made it clear just how closely staff link their perceptions about the change experience with the computer system. From their perspective the two are inextricably linked. System instability, as both the system and paperwork requirements underwent numerous modifications, during the first few months was unsettling for staff but also gave them an opportunity to influence and improve system functionality.

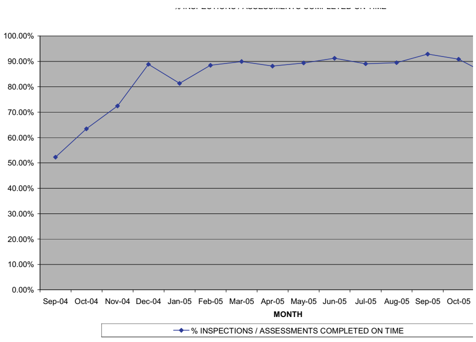
4.6 Data validation to enforce system compliance

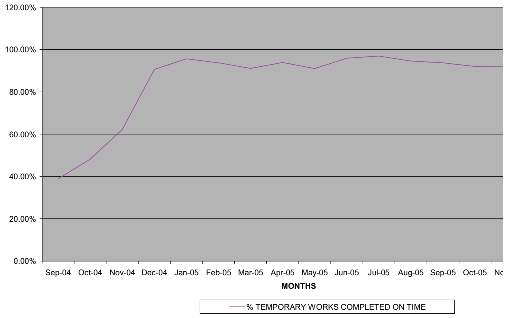
The collection of comprehensive maintenance data was essential to the asset management team as this information is critical for predictions of asset deterioration rates, repair costs, identification of locations where assets are failing abnormally etc. Operations Centre management and supervisors also required good data for budget and resource planning, efficient programming of works, measurement of compliance with the documented service levels, and compliance with the traceability requirements of Council's insurers. To make future data analysis possible, considerable effort went into developing codes for assets, activities, defects, failure causes etc. This was intended to simplify data collection on site without losing valuable information. However, field staff had a long history of recording simply the date they attended the site and a statement like "job done" on the printout of a customer request so collection of data as expected was overwhelming to many and others simply couldn't see the point.

When the system was first rolled out to staff, it did not incorporate any data validation. Analysis of data collected in the early days therefore revealed a clear case of 'garbage in equals garbage out'. Errors included blank fields, no risk assessment, inspection times later that the date the issue was repaired, no record of who undertook the inspection or repair, no record of how long the repair took, what plant or materials were used– none of which was surprising given that staff had never before been required to record their assessment of a site or the work they had done in such a structured manner. The introduction of data validation meant that they could not close off a request unless key data fields were completed with valid information. This made data analysis possible.

4.7 Performance reporting and auditing to drive improvement

Compliance with the strategic documents required a change in the way staff operated however the move toward risk based prioritisation of maintenance activities did not commence in earnest until the system could produce performance reports. This is evidenced in the graphs in Figure 9. The system was introduced in August 2004, performance reports became available toward the end of December 2004. These reports made key staff change their own ways and encourage others to follow. They realised that they would not be able to achieve performance targets if they continued with the old way of working - sticking to the old ways began to look unattractive. Performance reports therefore made it easier for managers to drive process improvements.





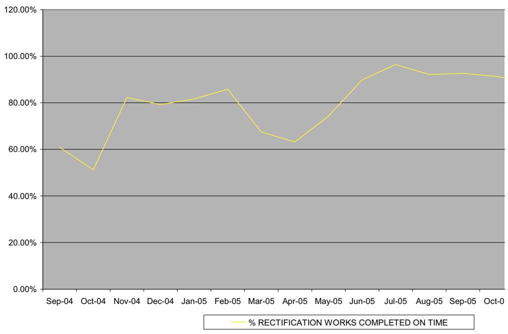


Figure 9 Compliance with service level timeframes

Improvement was seen almost immediately in terms of inspecting and assessing issues and making them safe however delivery of defect rectification target times was more difficult for staff to manage. This required systematic prioritisation of works. The large dip in performance in April resulted from poor management of footpath repair works generated by Council's new asset inspection program. The volume of footpaths requiring repair is still much greater than it had been before the introduction of inspections but now the staff interrogate the data to find up coming work and program it for repair.

Performance reporting was supplemented by performance audits undertaken under the direction of the Manager Operations. These audits considered all phases of the new work processes from completing and filing the necessary paper work, to site activities, risk assessments and data entry. Audits are undertaken regularly and non-conformance reports are prepared requiring officers to take corrective actions within a specified timeframe.

5 QUANTIFYING THE BENEFITS

Since the introduction of maintenance service levels and the supporting system in August 2004, the most obvious benefit has been a reduction in the number of requests made by customers for repair to a Council asset. This suggests that that the introduction of a proactive regular asset inspection program has enabled Council to identify and repair assets before the Community become sufficiently irritated by the issue to contact Council

Increased staff productivity is also evident in this graph as more issues are now able to be repaired without any significant increase in resources or funding, a direct benefit of programming the work efficiently. Data collected at the issue inspection phase includes scoping of the required maintenance activity including plant labour and material requirements making it possible for supervisors to program similar activities together according to location thereby reduce delays due to travel. Previously staff moved from one part of the municipality to another or returned frequently to the Operations Centre to collect required equipment and materials.

Figure 11 summarises Council's performance in audits undertaken by Local Government insurers Civic Mutual Plus. This graph of the scores provided show measurable improvement across all assets now managed in a more strategic and clearly auditable manner. These improvements have a direct impact in reducing Council's annual insurance premium.

Improved data collection also provides decision makers with a wealth of information that can be used to improve service delivery. Staff and plant utilisation can be measured; costs of specific maintenance activities can be interrogated. Decision makers can now identify the most likely cause of particular defects and identify and put in place improved measures to manage problem sites where there are multiple requests by customers, or where repaired defects reoccur suggesting underlying problems.

6 FUTURE CHALLENGES

Now that Operations Centre staff are implementing a strategic risk-based approach to asset maintenance delivering clear benefits to the community the challenge facing the asset management team is twofold. To keep the momentum going amongst staff to ensure they don't get pressured by other parts of the organisation into becoming reactive whilst spreading an understanding of strategic asset management principles to decision makers throughout the other phases of the asset lifecycle. Some work has already commenced.

Acknowledgments

Knox City Council Operations Centre Staff

System Developer JRA Pty Ltd

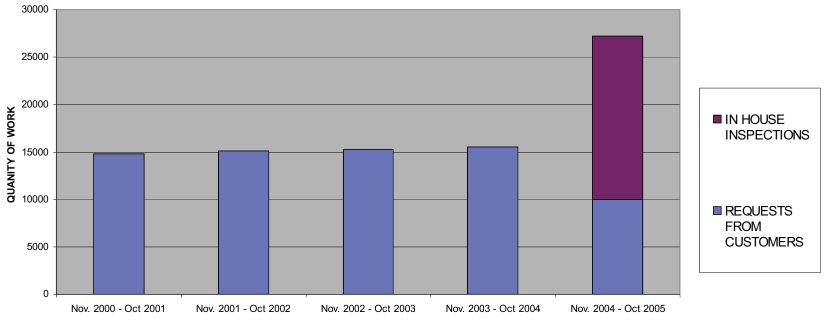


Figure 10 Source of maintenance activity

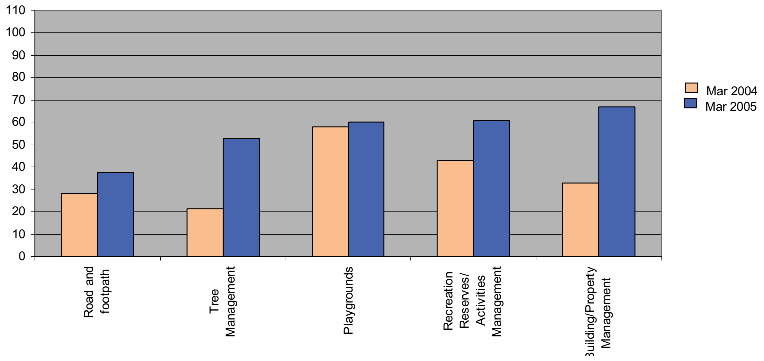


Figure 11 Liability insurance audit results

COMPUTATIONAL FEA MODEL OF A COUPLED PIEZOELECTRIC SENSOR AND PLATE STRUCTURE FOR ENERGY HARVESTING

M.F. Lumentut, K.K. Teh and I. Howard

Department of Mechanical Engineering

Curtin University of Technology

e-mail: I.Howard@curtin.edu.au

Abstract: his paper presents a mathematical model of a piezo-plate energy-harvesting scheme. An analytical method is used to generate a finite element model of the coupled piezoelectric sensor element using Love-Kirchhoff's plate theory. Constitutive equations for a single layer plate element are formulated. The polarisation of the piezoelectric sensor bounded on the upper plate structure is due to ambient vibration exerted on the structure. Forced vibration of the smart structure will create strain energy within the crystalline structure of the piezoelectric material. The resulting electric field generated by the sensor element was modelled using a linear thickness interpolation function and the meshed plate elements were modelled using four-node rectangular elements with three degrees of freedom for each node. The structural eigenmodes and dynamic response of the coupled piezo- plate system were solved by using modal analysis and Newmark-β integration methods respectively. The analysis is demonstrated with both dynamic displacement and electric voltage responses to an applied step force. Further modelling of the smart structure is aimed at maximising the power generation capability.

Keywords: energy harvesting, piezoelectric sensor, Love-Kirchhoff's plate, and Finite element analysis.

1 INTRODUCTION

The investigation of piezoelectric elements as smart structures has many wide-ranging applications in the engineering industry. The ability of piezoelectric materials to transform mechanical strain, induced by ambient vibration, to electrical potential is well known, as evidenced in the direct mode by using the material as a sensor. On the other hand, piezoelectric material which has the ability to transform electrical field to mechanical strain is also well-known, as the converse mode of using the material as an actuator. Numerical analysis of these smart structures using a three dimensional finite element model was first presented by Allike and Hughes in 1970[1]. However, a three dimensional finite element analysis of the integrated piezo and plate structure was challenging and generally demanded large computational requirements. It was realized that the piezoelectric patches were too thin and two dimensional modeling of these patches was sufficient. Other numerical models using laminated rectangular plates with piezoelectric material bonded on the surface have been previously developed[2,3,4], with other researchers using multilayer composite plates[5,6]. Several publications,[7,8,9] have reported the use of numerical models to simulate the dynamic response of the coupled elastic and electric field of the smart structures.

The previous scenarios, mostly, discussed static and dynamic systems using the piezoelectric sensor and actuator bounded on the structures. Moreover, recent developments of piezoelectric technology involve the usage of sensors patched on the structure subject to ambient mechanical vibration. The induced strain energy can be converted to useful electrical energy capable of being stored on electrical devices such as batteries or capacitors. Such a technique is generally referred to as the energy harvesting technique. The development of energy harvesting techniques for powering smart structures and embedded sensors has received increased attention over the past decade[10]. Analytical and experimental analyses have been used to investigate power harvesting from PZT elements to power electric devices or for recharging of batteries. Embedded PZT materials in a vibrating machine environment can be used as the required power source provided the vibration source does not stop. A more useful development would be for the PZT element to store its energy into a rechargeable battery for later use. Further analytical and experimental studies with optimising power flow for adaptive energy harvesting and self-power harvesting have been developed [11,12,13,14].

The main objective of this paper is to present a mathematical model of an energy harvesting technique with bonded piezo-sensors on a plate element. Mechanical strain energy induces polarization in the piezo-sensor thus creating an electric field. The mathematical model introduces a four-node non-conformed quadrilateral element with a total of twelve nodal degrees of freedom, formulated using Love-Kirchhoff's plate theory. The paper presents results from a two-sided fixed-fixed plate using numerical algorithms solved using MATLAB, based on the suggested formulation.

1.1 CONSTITUTIVE EQUATION OF PLATE ELEMENT

Classical plate theory can be used to derive the equations of motion for plates by assuming that the shear deformation effect is negligible. The displacement vectors u,v, and w of an arbitrary point in the deformed element can be written in terms of the mid-surface of the plate, as described in Figure 1a and Figure 1b and equations 1, 2 & 3.

$$u = -z\frac{\partial w}{\partial x},\quad (1)$$

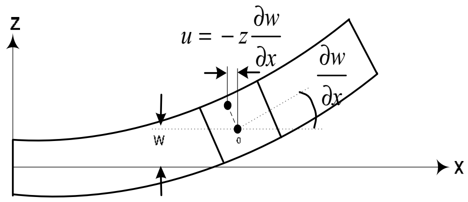


Fig. 1a. The deformations of plate element with respect to x axis

$$v = -z\frac{\partial w}{\partial y},\quad (2)$$

$$w = w(x,y),\quad (3)$$

The corresponding strain-displacement vector relationship in the x and y directions, can be formulated as,

$$\{\varepsilon\} = \begin{Bmatrix} \varepsilon_x \\ \varepsilon_y \\ \gamma_{xy} \end{Bmatrix} = -z \begin{bmatrix} \frac{\partial^2 w}{\partial x^2} \\ \frac{\partial^2 w}{\partial y^2} \\ 2\frac{\partial^2 w}{\partial x \partial y} \end{bmatrix} .(4)$$

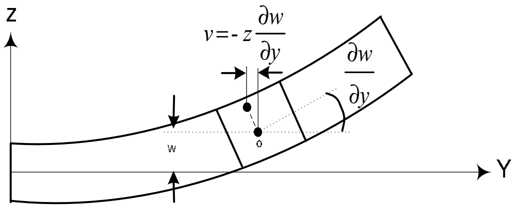


Fig. 1b. The deformation of plate element with respect to y axis.

A state of plane stress is assumed and the corresponding stress-strain relationship is,

$$\{\sigma\} = \frac{C}{1-v^2}\begin{bmatrix} 1 & v & 0 \\ v & 1 & 0 \\ 0 & 0 & \frac{1-v}{2} \end{bmatrix} \begin{Bmatrix} \varepsilon_x \\ \varepsilon_y \\ \gamma_{xy} \end{Bmatrix}, \quad (5)$$

Or

$$\{\sigma\} = [D_m]\{\varepsilon\} . \quad (6)$$

1.2 CONSTITUTIVE EQUATION OF PIEZO-SENSOR

The piezoelectric-sensor has a direct mode electrically on its material due to the exerted mechanical strain on the structure. The matrix equation relating the mechanical and electrical quantities can be written as [15],

$$\{D\} = [e]\{\varepsilon\} + [\xi]\{E\}, \quad (7)$$

where $\{D\}$, $[e]=[d][D_m]$, $\{\varepsilon\}$, $[\xi]$, $[d]$, $[D_m]$ and $\{E\}$ are the electric displacement vector, piezoelectric stress coefficient, strain vector, dielectric matrix at constant strain, piezoelectric constant, stiffness coefficient and electric field vector respectively. In this case, the converse mode gained by the actuator is neglected.

Discritised electric field {E} induced by ambient vibration generates polarisation in the piezo-sensor material, in the z-direction along the sensor plate thickness. The subsequent electrical potential is assumed linear and is formulated as,

$$\{\delta_{(z)}\} = [\Phi^S_{(z)}]\{\delta^s_{(z)}\}, \quad (8)$$

where $\Phi^{s}_{(z)} = \dfrac{z - \dfrac{t_p}{2}}{t_s}$ is the shape function over the interval tp/2 < z < tp/2 + ts.

The electrical field is a function of the electrical potential with negative gradient operator,

$$\{E\} = -\nabla\left\{\delta_{(z)}\right\} = -\left[\Psi^{s}_{(z)}\right]\left\{\delta^{s}_{(z)}\right\} \tag{9}$$

where ∇ is a gradient operator, first derivative of the shape function with respect to thickness direction, giving,

$$\left[\Psi^{s}_{(z)}\right] = \begin{bmatrix} 0 & 0 & \dfrac{d\Phi^{s}_{(z)}}{dz} \end{bmatrix} = \begin{bmatrix} 0 & 0 & \dfrac{1}{t_s} \end{bmatrix}. \tag{10}$$

2 FINITE ELEMENT FORMULATION

2.1 Discretized Element Plate Matrix

As mentioned previously, the classical plate theory can be further dealt with by establishing a 12-term polynomial function to model the non-conforming transverse displacement, w(x,y). The three degrees of freedom at each node wn(x,y), θxn and θyn can then be formulated as shown in Figure 2.

The nodal degrees of freedom for the four-noded rectangular element, can be expressed in the vector form, $\{\delta^{p}_{n(x,y)}\} \in \{\delta^{p}_{i(x,y)}, \delta^{p}_{j(x,y)}, \delta^{p}_{k(x,y)}, \delta^{p}_{l(x,y)}\}$ and $\forall$ {wn, θxn, θyn}. All variables of displacement obtained will be substituted into a polynomial function where the results will give the consistent displacement function as,

$$w(x,y) = [\Phi^{p}_{n(x,y)}]\{\delta^{p}_{n(x,y)}\} \tag{11}$$

where $[\Phi^{p}_{n(x,y)}]$ is the shape function of the displacement field for the non-conforming plate element with reference to transversal displacement function,

$$[\Phi^{p}_{n(x,y)}] = [\zeta][\zeta^{p}_{n(x,y)}]^{-1}, \qquad n \in \{i,j,k,l\}. \tag{12}$$

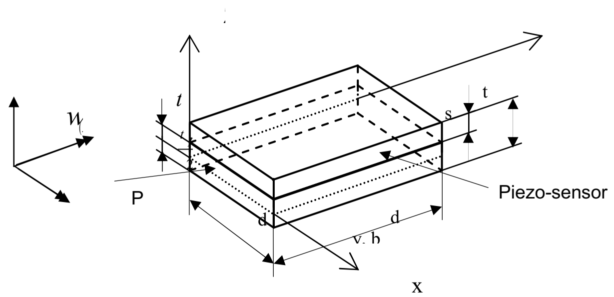


Fig.2. Geometry element of the coupled piezo-sensor and plate element.

The strain-displacement relationship from equation (4) can be expanded by forming the second order partial differential equation of shape function, equation (12), with respect to x and y axes of the element plate in each joint as follows,

$$\varepsilon=-z\begin{bmatrix} \frac{\partial^2[\Phi^p_{i(x,y)}]}{\partial x^2} & \frac{\partial^2[\Phi^p_{j(x,y)}]}{\partial x^2} & \frac{\partial^2[\Phi^p_{k(x,y)}]}{\partial x^2} & \frac{\partial^2[\Phi^p_{l(x,y)}]}{\partial x^2} \\ \frac{\partial^2[\Phi^p_{i(x,y)}]}{\partial y^2} & \frac{\partial^2[\Phi^p_{j(x,y)}]}{\partial y^2} & \frac{\partial^2[\Phi^p_{k(x,y)}]}{\partial y^2} & \frac{\partial^2[\Phi^p_{l(x,y)}]}{\partial y^2} \\ 2\frac{\partial^2[\Phi^p_{i(x,y)}]}{\partial x\partial y} & 2\frac{\partial^2[\Phi^p_{j(x,y)}]}{\partial x\partial y} & 2\frac{\partial^2[\Phi^p_{k(x,y)}]}{\partial x\partial y} & 2\frac{\partial^2[\Phi^p_{l(x,y)}]}{\partial x\partial y} \end{bmatrix}\{\delta^p_{n(x,y)}\} \quad . \tag{13}$$

Equation (13) can be simply rewritten as follows,

$$\varepsilon = z\left[\Psi^p_{n(x,y)}\right]\left\{\delta^p_{n(x,y)}\right\}, \quad n \in \{i,j,k,l\}. \tag{14}$$

Governing Differential Dynamic Equations

By invoking Hamilton's principle for a conservative dynamic system and applying Lagrange's simplification, the general dynamic equation can be expressed as,

$$\sum_{\eta=1}^{s}\left[\frac{d}{dt}\frac{\partial \Pi}{\partial \dot{u}_\eta}-\frac{\partial \Pi}{\partial u_\eta}-F_\eta\right]=0, \tag{15}$$

where,

$$\forall \ \Pi = KE - PE + PEE,$$

$$\eta \in \left\{\left\{\dot{\delta}^p_{n(x,y)}\right\}\left\{\delta^p_{n(x,y)}\right\},\left(\delta^s_{(z)}\right)\right\} ; F\eta \in \{\{Fs\},\{Fc\},\{q\}\} .$$

The kinetic energy term is given by,

$$KE = \frac{1}{2}\iiint \rho_p \left\{\dot{w}_{(x,y)}\right\}^T \left\{\dot{w}_{(x,y)}\right\} d(vol) + \frac{1}{2}\iiint \rho_s \left\{\dot{w}_{(x,y)}\right\}^T \left\{\dot{w}_{(x,y)}\right\} d(vol), \tag{16}$$

and the potential energy can be written,

$$PE = \frac{1}{2}\iiint \{\varepsilon\}^T \{\sigma\} d(vol) \quad . \tag{17}$$

The electrical energy term can be formed as,

$$PEE = \frac{1}{2}\iiint \{E\}^T \{D\} d(vol) \quad . \tag{18}$$

Work done on the system is due to the surface force, concentrated force and electrical charge density and can be written as,

$$\Omega = \iint \{\delta_e\}^T \{Fs\} d(area) + \{\delta_e\}^T \{Fc\} - \iint \left\{\delta_{(z)}\right\}^T \{q\} d(area) \tag{19}$$

The solution procedure involves substituting equations (6), (7), (9) and (14) into related expressions of equations (16) to (19) and differentiating with respect to all displacement vector parameters of the coupled sensor and plate structure. As given in equation (14), the resulting non-homogenous matrix differential equation for the smart structure can be formed as,

$$\begin{bmatrix} [M^{(ps)}] & 0 \\ 0 & 0 \end{bmatrix} \begin{Bmatrix} \ddot{\delta}^{p}_{n(x,y)} \\ \ddot{\delta}^{s}_{(z)} \end{Bmatrix} + \begin{bmatrix} [K_{\phi\phi}] & [K_{\phi\theta}] \\ [K_{\theta\phi}]^T & [K_{\theta\theta}] \end{bmatrix} \begin{Bmatrix} \delta^{p}_{n(x,y)} \\ \delta^{s}_{(z)} \end{Bmatrix} = \begin{Bmatrix} F_{\phi} \\ F_{\theta} \end{Bmatrix}, \tag{20}$$

where:

$$[\mathrm{M(ps)}] = \sum_{i=1} \left(\iiint \rho_i \left[\Phi^{p}_{n(x,y)} \right]^T \left[\Phi^{p}_{n(x,y)} \right] d(\mathrm{vol})_i \right), \tag{21}$$

$$[\mathrm{K\Phi\Phi}] = \sum_{i=1} \left(z^2 \iiint \left[\Psi^{p}_{n(x,y)} \right]^T \left[D_i \right] \left[\Psi^{p}_{n(x,y)} \right] d(\mathrm{vol})_i \right) \tag{22}$$

$$[\mathrm{K\Phi\theta}] = [\mathrm{K\theta\Phi}]\mathrm{T} = -\sum_{s=1} \left(\frac{z}{2} \iiint \left[\Psi^{p}_{n(x,y)} \right]^T \left[e \right]^T \left[\Psi^{s}_{(z)} \right] d(\mathrm{vol})_s \right), \tag{23}$$

$$[\mathrm{K\theta\theta}] = -\sum_{s=1} \left(\iiint \left[\Psi^{s}_{(z)} \right]^T \left[\xi_s \right] \left[\Psi^{s}_{(z)} \right] d(\mathrm{vol})_s \right), \tag{24}$$

$$\{\mathrm{F\Phi}\} = \sum_{i=1} \left(\iint \left[\Phi^{p}_{n(x,y)} \right]^T \{Fs_i\} d(area)_i \right) + \left[\Phi^{p}_{n(x,y)} \right]^T \{Fc_i\}, \tag{25}$$

$$\{\mathrm{F\theta}\} = -\sum_{i=1} \left(\iint \left[\Phi^{s}_{(z)} \right]^T (q_i) d(area)_i \right). \tag{26}$$

To obtain the eigenmodes, dynamic displacement and electric voltage, the matrix differential equation (20) can be solved using Guyan Reduction to separate the degrees of freedom of the variable displacements where the equations yielded reflect more appropriately the independent equations for electric voltage and linear dynamic displacement. At this point, the Newmark-β method was used to solve for the dynamic response of the piezo-plate vibration system.

2.2 Application of Boundary Condition

Equations (21) to (26) can be solved by using integral algebra and by incorporating appropriate geometric boundary conditions. In view of the different geometrical boundary conditions for the sensors and the base plate, special attention must be taken when modifying equations (21) to (24) to reflect these conditions.

3 RESULT AND DISCUSSION

3.1 Numerical Example

A MATLAB program was written based on the proposed theory to predict eigenmodes, dynamic displacement, and electrical voltages under transient response conditions. The characteristics for the piezo-sensor using PZT PSI-5A4E (Piezo Systems, INC) and the plate are listed in Table 1 as follows.

Table 1. Characteristic data

Item	Piezo-sensor	Plate
C (GPA)	66	207
υ	0.3	0.3
ρ (kg/m3)	7800	7870
t (m)	0.000267	0.002
d31 (m/V)	-190e-12	-
d33 (m/V)	390e-12	-
ξ_{33} (F/m)	1.602e-8	-
l x w (m2)	0.25 x 0.2	1 x 0.8

The example used a two-sided fixed-fixed rectangular plate with piezo-sensors bonded on the upper surface as shown in Figure 3. The plate was subjected to an externally applied dynamic step force of 1 Newton at node 2, at time t=0 sec.

Figure 3 shows the mesh pattern for the piezo-plate structure, consisting of 16 elements (4x4) and 16 elements (4x4) of the piezo-sensors. The Piezo-sensor was bonded over the upper plate surface. The numerical methods outlined above were used to solve the resulting finite element model using the Newmark-β method with the application of the MATLAB program.

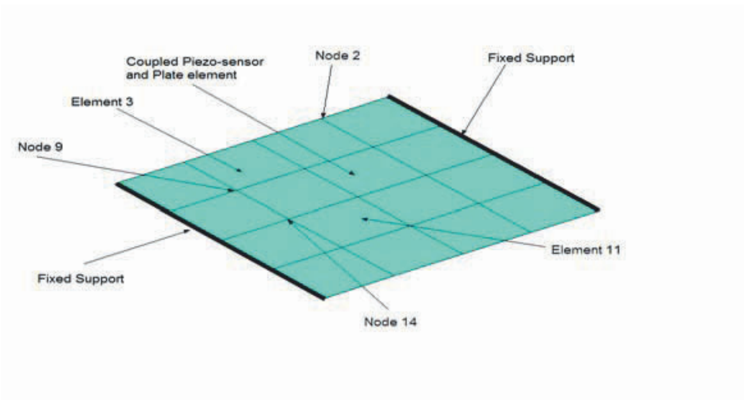


Fig. 3. Geometry of the coupled piezo-sensor and plate element.

3.2 Eigenmodes

Figures 4a to 4d show the eigenvectors corresponding to the first four natural frequencies of the composite plate-piezo structure. The eigenvector corresponding to the natural frequency of 11.22 Hz shown in Figure 4a indicates clearly the first transverse vibration of a fixed-fixed plate. Figure 4b shows the first twisting vibration mode at 14.31 Hz. The next two eigenvectors shown in Figures 4c and 4d indicate the second bending and twisting modes at 27.20 Hz and 31.23 Hz respectively. To optimize the location of the piezo patch on the plate, modal analysis can be used to understand the prediction of deflected shapes at each resonance frequency. The generated electric voltage signal from the sensors can be shown as a function of time, resulting from the transient displacement response of the piezo patch.

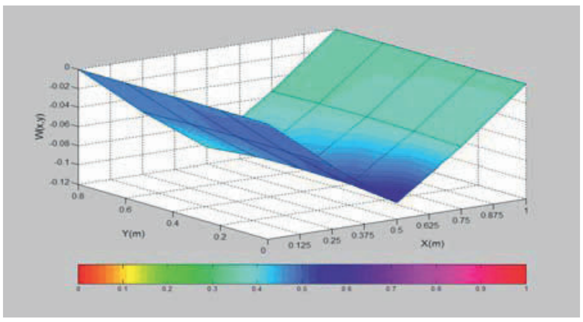


Fig. 4a. First eigenmode at ωn = 11.22 Hz

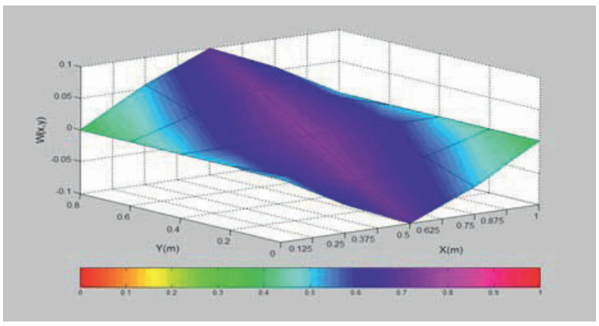


Fig.4b. Second eigenmode at ωn = 14.31 Hz.

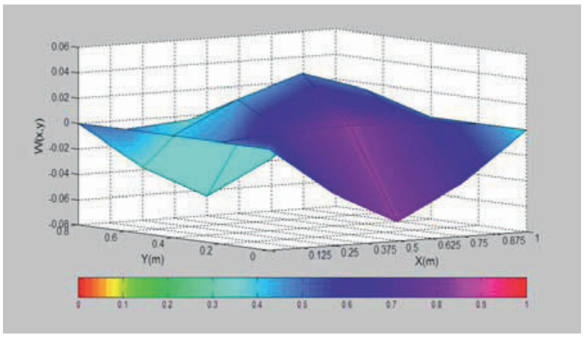


Fig.4c. Third eigenmode at ωn = 27.20 Hz

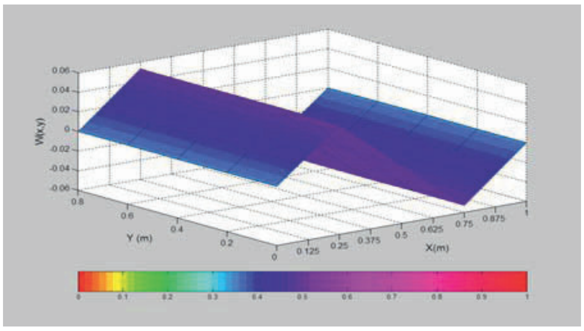


Fig.4d. Fourth eigenmode at ωn = 31.23 Hz.

3.3 Dynamic Displacement History

To illustrate the numerical method, a step force was used as an input force. As shown in Figure 3, a step force of 1N was applied to the plate at node 2 at t=0 sec. The resulting transient dynamic displacement of the plate is shown in figure 5a and figure 5b giving the resultant dynamic displacement at node 9 and node 14 respectively. To reduce the computational requirement, the number of elements in the FE model was reduced to the initial 4x4 system as shown in Figure 3. For the results shown, the sampling frequency for the simulation was chosen to be 1646 Hz. To avoid aliasing of the time domain simulation, the time history was sampled at least twice the highest eigen-frequency present in the system. Each node of the plate-piezo structure will result in a different dynamic response. Figures 5a and 5b show the response obtained at two nodes over a one second time period. Initially the simulation covered 10 seconds, to provide a fine frequency response of the structure. As evidenced from the dynamic displacement, several frequencies are present in the time waveform, with the most predominant having a frequency close to 14 Hz.

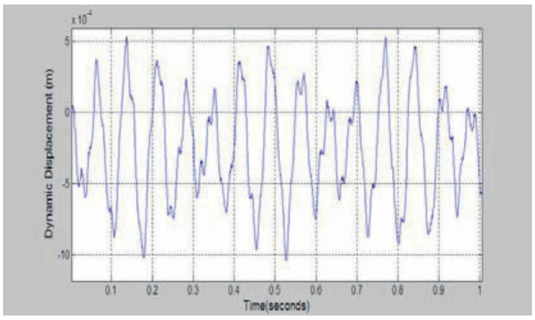


Fig. 5a. Dynamic displacement at node 9.

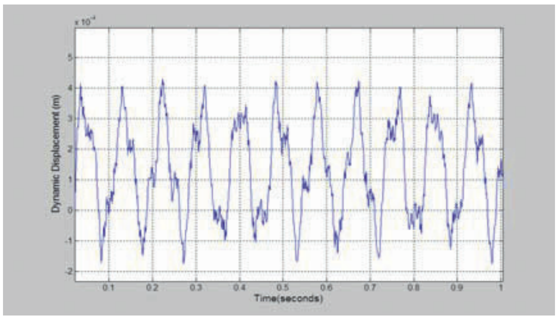
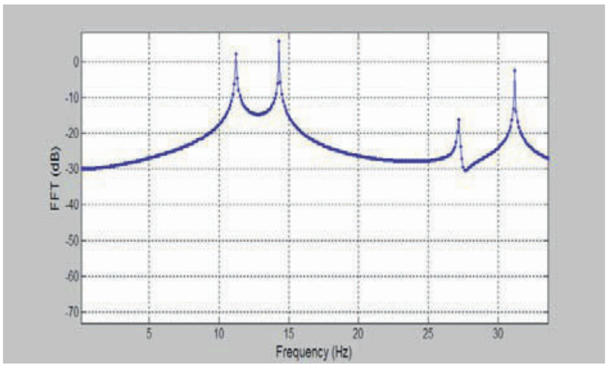


Fig. 5b. Dynamic displacement at node 14.

3.4 Frequency Response and the Transfer Function

The Fast Fourier Transform (FFT) was used to analyse the dynamic displacement and obtain an estimate of the frequency response of the structure as a result of the applied step force. Figure 6a shows the frequency response calculated from the dynamic displacement in Figure 5a (node 9) over the frequency range from 0-35Hz. It should be noted that the step force excited several resonances of the system over this frequency range. Closer inspection shows that these resonances are the same as those predicted from the eigenmodes as shown in Figure 4a-4d. This was to be expected as the abrupt step force would be expected to excite the resonances of the structure. The frequency response was scaled as magnitude and plotted on a dB scale.

The Transfer Function reflects the characteristic dynamic frequency and the relationship with the eigenmodes and eigenvectors of the coupled piezo-sensor and plate structure. The Transfer function refers to a specific dynamic input force and an output measurement (displacement or acceleration) of the structure. Both the FFT and the Transfer function have unique behaviour based on the chosen location for the input force and the output measurement location. Figure 6b shows the transfer function of the system calculated for the combined piezo-plate structure between nodes 2 and 9.



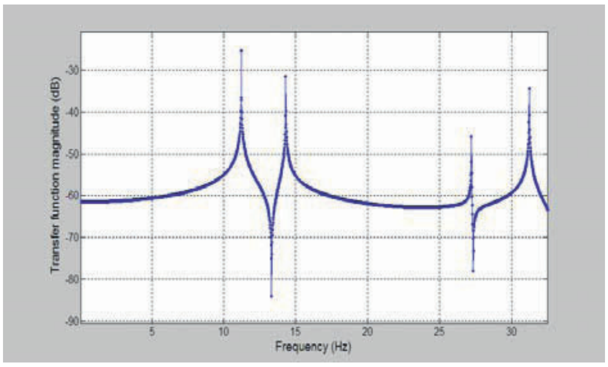


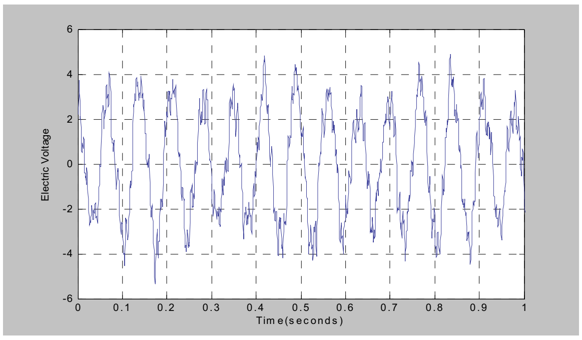
Fig. 6a. Frequency response obtained from node 9. **Fig. 6b. Transfer Function between nodes 2 and 9**

The transfer function result clearly shows the first four natural frequencies as dominant peaks, at the same frequencies as calculated for the eigenmodes shown in Figure 4 corresponding to 11.22 Hz, 14.31 Hz, 27.20 Hz and 31.23 Hz respectively. These resonances are also identical with those obtained from the frequency response as given in Figure 6a. This is further evidence of the correct simulation result obtained by the numerical method.

The theoretical model used in this paper neglects any damping effects present in the plate or piezo material. The overall heights of the transfer function magnitudes at resonance shown here are thus dependant on the frequency spacing chosen for the analysis, as the actual resonant peak in the transfer function will have infinite height. The transfer function and FFT show the direct response of the piezo-sensor bonded on the plate structure for a particular force input and dynamic response output location. The magnitudes obtained by both FFT and transfer function can be used for optimization studies to identify the best overall location of sensor patches to provide maximum piezo power output for a given frequency of input. This would require further development of the transfer function to provide the transfer function relating the input dynamic force location to the output piezo strain. For power generation optimisation studies, the FFT and transfer function relating the ambient vibration input conditions to the piezo output electric field could also be used.

3.5 Electric Voltage Time History

To further illustrate the numerical approach, the electric voltage time history resulting from the force step input of 1N, located mid-span on the fixed-fixed support of the structure shown in Figure 3 was produced. The characteristics of the electrical voltage refer to the existences of dynamic input from ambient vibration exerted of the plate where sensors are bonded on the upper plate surface. Figures 7a and 7b show the subsequent voltage time history (with zero mean) over a short time length from elements 3 and 11 respectively.



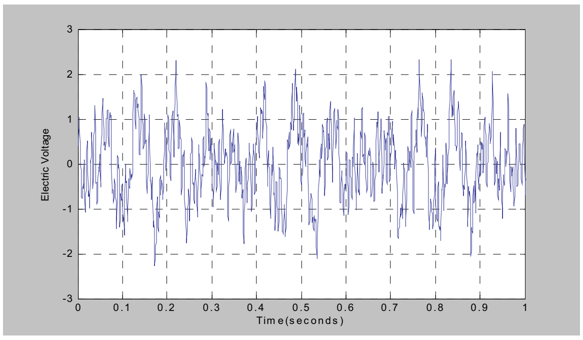


Fig. 7a. Electric Voltage at element 3 **Fig 7b. Electric Voltage at element 11**

It should be noted that the plate was initially at rest, and the input force has excited at short time interval of the structure. In this case, two elements show the characteristics of electric voltage where these depend on the strain of the elements due to the dynamic displacement. The effects of the strain generate the electric voltage at each time step of the transient response. As expected, the voltage time history contains frequency components at the structural resonance frequencies. The next step in this research will involve experimental confirmation of these results and design optimisation to maximise the power output under free and forced vibration conditions.

4 CONCLUSION

This paper has presented a finite element model of a smart structure for energy harvesting. The constitutive equations have been formulated using the Love-Kirchhoff's plate theory and piezo-sensor theory. An example was used to demonstrate the validity of the mathematical model. The dynamic equations obtained in the model have been calculated using Guyan reduction and the Newmark Beta algorithm. Acceptable results for dynamic displacement, electric potential, eigenvalues and their corresponding eigenvectors, frequency response and the transfer function have been obtained and discussed. Ongoing research is being conducted to compare the versatility of the current model with other existing models and with experimental work.

NOMENCLATURE		Greek Letters	
u	displacement of plate element in x direction, m	ε	strain vectors
v	displacement of plate element in y direction, m	σ	stress vectors, N/m2
w	displacement of plate element in z direction, m	ξ	dielectric matrix at constant strain, F/m
C	modulus of elastic, N/m2	υ	poison ratio
D	electric displacement vectors, C/m2	Φ	shape function
d	piezoelectric constant, m/V	$\delta_{(z)}$	electric potential, Volts
e	piezoelectric stress coefficient, C/m2	δ	nodal vectors, m
E	electric field, V/m	Ψ	differential operator of shape function
Dm	stiffness coefficient, N/m2	ρp	density of plate, kg/m3
t	thickness, m	ρs	density of sensor, kg/m3
l	length, m	Π	total of energy, Nm
w	width, m	Ω	work done of the system, J
F	force, N		
KE	Kinetic Energy, J		
PE	Potential Energy, J		
PEE	electrical energy, J		
K	stiffness matrix, N/m		
M	consistent-mass matrix, kg		
Subscripts		**Superscripts**	
x	relative to x-axis	s	refer to sensor element
y	relative to x-axis	p	refer to plate element
z	relative to z-axis	ps	refer to plate-sensor element
n	refer to joints of plate element	T	matrix transpose
θθ	refer to plate		
Φθ	refer to sensor–plate		
θΦ	refer to plate-sensor		
Φ	refer to sensor		
θ	refer to plate		

5 REFERENCES

[1] Allik, H & Hughes, T.J.R. (1970). "Finite Element Method for Piezoelectric Vibration." International Journal for Numerical Methods in Engineering, Vol. 2, pp. 151-157.

[2] Moita, J. M., Correia, I.F.P. & Soares, C.M.M (2004). "Active Control of Adaptive Laminated Structures with Bounded Piezoelectric Sensors and Actuators." Computer & Structures, Vol. 82, pp. 1349-1358.

[3] Tzou, H. S. & Tzeng, C.I. (1990). "Distributed Piezoelectric Sensor/Actuator Design for Dynamic Measurement/Control of Distributed Parameter System: A Piezoelectric Finite Element Approach." Journal of Sound and Vibration Vol. 138, No. 1, pp. 17-34.

[4] Fernandes, A. & Pouget, J. (2003). "Analytical and Numerical Approach to Piezoelectric Bimorph." International Journal of Solid and Structures, Vol. 40, pp. 4331-4352.

[5] Saravanos, D. A., Heyliger, P.R. & Hopkins, D.A. (1996). "Layerwise Mechanics and Finite Element for the Dynamics Analysis of Piezoelectric Composite Plate." International Journal of Solids Structures, Vol. 34, No.3, pp. 359-378.

[6] Kogl, M. & Bucalem, M.L. (2005). " Analysis of Smart Laminates Using Piezoelectric MITC Plate and Shell Elements." Computer and Structures, Vol. 83, pp. 1153-1163.

[7] Sekouri, E. M., Hu, Y. & Ngo, A.D. (2004). "Modeling of a Circular Plate with Piezoelectric Actuators." Mechatronics, Vol. 14, pp. 1007-1020.

[8] Liu, X., Wang, Q., Quek, S.T. (2002). "Analytical Solution for Free Vibration of Piezoelectric Coupled Moderately Thick Circular Plates." International Journal of Solids and Structures, Vol. 39, pp. 2129-2151.

[9] Taciroglu, E., Liu, C.W., Dong, S.B. & Chun, C.K. (2004). "Analysis of Laminated Piezoelectric Circular Cylinders under Axisymmetric Mechanical and Electrical Loads with a Semi-analytical Finite Element Method." International Journal of Solids and Structures Vol. 41, pp.5185-5208.

[10] Sodano, H. A., Inman,D.J. & Park, G. (2004). "A Review of Power Harvesting from Vibration Using Piezoelectric Materials." The Shock and Vibration Digest Vol. 36, No. 3, pp. 197-205.

[11] Ottman, G. K., Hofmann,H.F., Bhatt, A.C. & Lesieutre, G.A. (2002). "Adaptive Piezoelectric Energy Harvesting Circuit for Wireless Remote Power Supply." IEEE Transactions on Power Electronics, Vol.17, No.5, pp. 669-676.

[12] James, E. P., Tudor, M.J., Beeby, S.P., Harris, N.R., Glynne-Jones, P., Ross, J.N. & White, N.M. (2004). "An Investigation of Self-Powered System for Condition Monitoring Applications." Sensors and Actuators: A Physics ,Vol. 110, pp. 171-176.

[13] Roundy, S., Wright, P.K. & Rabaey, J. (2003). "A Study of Low Level Vibrations as a Power Sources for Wireless Sensor Nodes." Computer Communications ,Vol. 26, pp. 1131-1144.

[14] Glynne-Jones, P., Tudor, M.J., Beeby, S.P. & White, N.M. (2004). "An Electromagnetic, Vibration-Powered Generator for Intelligent Sensor System." Sensors and Actuators: A Physics, Vol. 110, pp. 344-349.

[15] Naillon, M., Coursant R.H. & Besner, F. (1983). "Analysis of Piezoelectric Structures by a Finite Element Method."ACTA Electronica, Vol. 25, No. 4, pp. 341-362.

STRATEGIES FOR PHYSICAL ASSET MANAGEMENT IN KHNP

W. Lee [a], S. Moh [b], S. Min [c] and B. Yang [d]

[a, b, and c] *Maintenance Planning & Engineering Department, Korea Hydro & Nuclear Power Co., 167 Samseong-dong, Gangnam-gu, Seoul, Korea.*

[d] *Mechanical Engineering, Pukyong University, 599-1 Daeyon3-dong, Nam-gu, Busan, Korea.*

Abstract: Strategic asset management (SAM) can be simply described by two key words, which are Asset Evaluation and Asset Enhancement. Asset value herein does not imply installation cost but rather refers to a market price of the asset when sold in a competitive market. Asset evaluation is a part of business administration field and there are many verified methodologies and practical examples. Therefore this paper would rather be focused on asset enhancement. Applied to physical assets, to increase current asset value and to decrease total cost of ownership (TCO), asset management should be concentrated upon graded approaches, proactive maintenance, and long-term strategies, three key topics of this paper.

Key Words: SAM, Asset Management, Graded Approach, Proactive Maintenance, Long-term Strategy, KHNP

1 INTRODUCTION

Korea Hydro and Nuclear Power Corporation (KHNP) has introduced concept of SAM into the physical asset management processes (See Figure 1) [1,2,3,4,5,6,7]. It consists of several processes including Functional Importance Determination (FID), Performance Management (PM), Corrective Action Program (CAP), Maintenance Optimization (MO), Functional Equipment Group (FEG), Work Management (WM), Investment Management (IM), and Inventory Optimization (IO). FID process is used to classify functional characteristics of equipment and to support graded approach philosophy through preparing critical inputs to other processes. PM, CAP, MO, and WM processes allow efficient resource allocation on proactive tasks and to conduct planned maintenance prior to components' functional failures. To bring long-term strategic concept into the physical asset management, KHNP has been implementing IM and IO processes. IM process includes automated appropriation request and evaluation as well as prioritizing tools for investment decision making, and Long Term Planning. The IO process is used to define appropriate safety stock for each important material and to track obsolescence issues.

2 GRADED APPROACH

Functional Importance Determination (FID) process is used to classify functional characteristics of each component to support graded approach concept in the physical asset management. The functional characteristics of each component consist of functional importance, component type, duty cycle, and service condition (See table 1) [8.9.10].

Functional importance is determined by answering 30 questions which cover considerations for nuclear safety, power production, and regulatory and environmental effects in case the component fail to perform its intended function(s). First 20 questions target critical component determination while the other 10 questions target non-critical component determination. If the component's functional failure does not result in any one of 30 questioned effects, it is determined as no impact. No impact components are considered as Run to Failure (RTF) components on which no preventive maintenance tasks are applied and only corrective maintenance tasks are performed when failed. The Probabilistic Safety Analysis (PSA) and the Maintenance Rule (MR) evaluation results are used to define criticality in nuclear safety [11]. Functional importance is widely used in almost all asset management processes as an input for decision making.

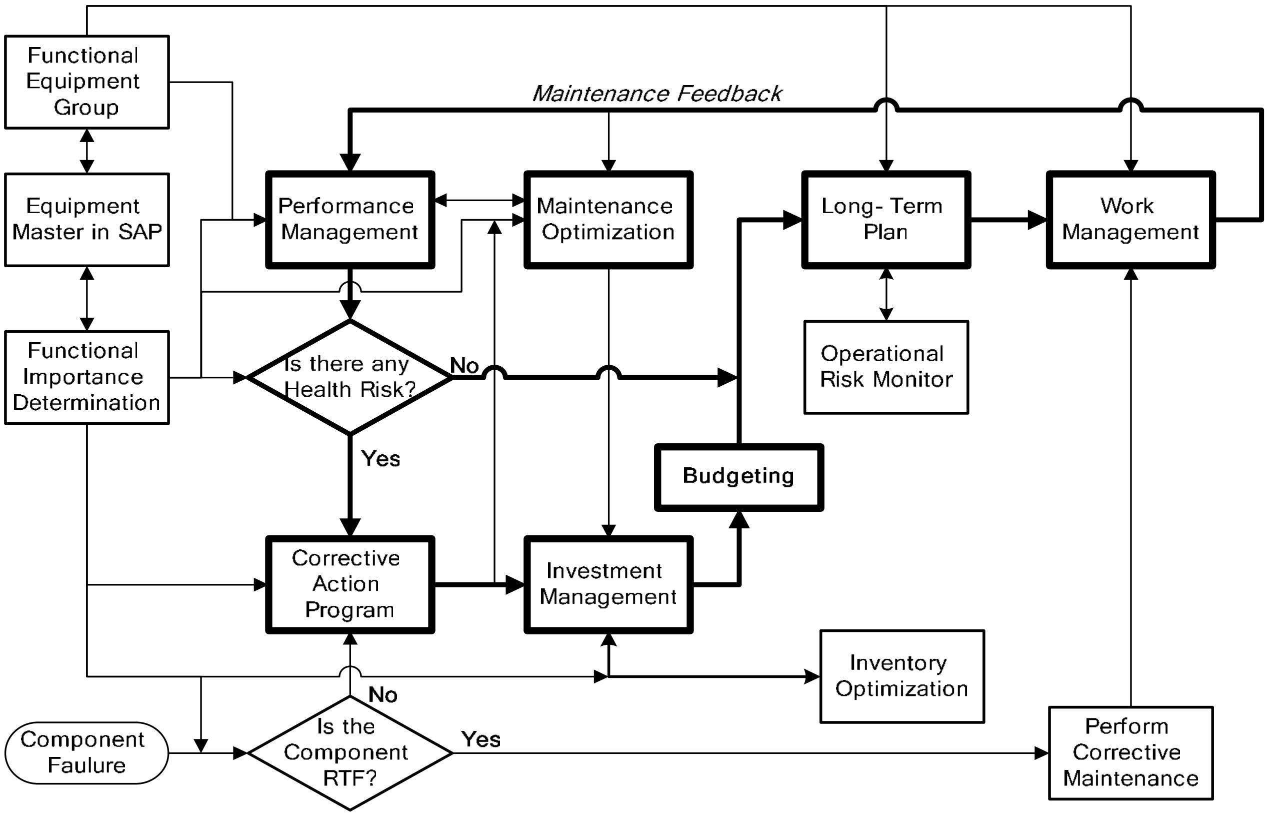


Figure 1 KHNP SAM process overview

Table 1

Components' functional characteristics determined through FID process

Characteristic	Code	Description	Comment
Functional Importance	A	Critical A	Both are fallen under "Critical" in PM program
	B	Critical B	
	C	Minor	Non-critical
	X	No Impact	RTF in PM program
Component Type	Char(4)	217 types	One type is selected
Duty Cycle	H	High	Defined by definitions developed in MO process for each component type.
	L	Low	
Service Condition	S	Severe	
	M	Mild	

If evaluated component is identified as "critical" or "minor," the component is included in scope of preventive maintenance program. We should subsequently define component type by selecting one from 217 types in terms of maintenance program since the selected component type is used directly to establish preventive maintenance program for the component in MO process. Once the component type is selected, we can define duty cycle and service condition for the component based on definitions developed in MO process for each component type. Both duty cycle and service condition are drivers for the interval of preventive maintenance task in MO process. Figure 2 shows the flowchart for the main steps of FID process.

3 PROACTIVE MAINTENANCE

Functional Equipment Group (FEG) process is used to develop FEGs which are tools used in the Work Management process and serves its purpose to electronically group works. In other words, we can easily and automatically extract work packages using FEGs [12,13,14].

A FEG is a group of equipment pieces which perform or support a certain function. Therefore a FEG boundary should be drawn based on the affecting border of a specific function. Figure 3 represents FEG development process and figure 4 shows an example of FEG boundary drawing.

The FEG process ensures consistent and accurate functional importance determination of components. After FEGs development, the functional importance of FEG is defined through FID process and the functional importance of related components can be simply defined based on the contributions to FEG's function.

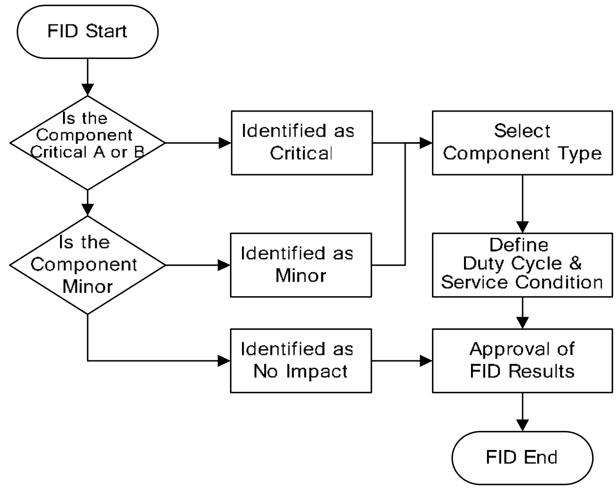


Figure 2 Overview of the FID process

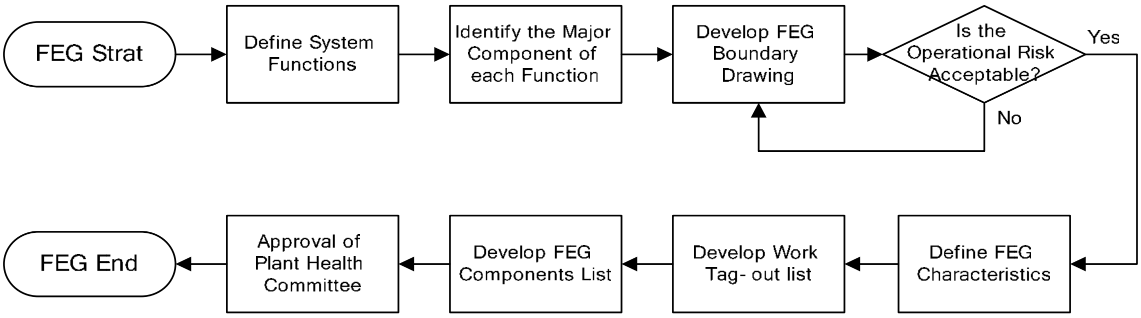


Figure 3 Simplified FEG process

Adapting FEGs into Work Management process, we can minimize the number of times important functions are removed from service thus optimize unavailability of these functions. Since FEGs are used as the basis in the Long Term Planning (LTP) and plant maintenance process, the application of FEGs reduces followings:

- The time required for schedulers to develop candidate work lists.
- Work load of Operations, Work Management and Maintenance.
- Potential for performance of work related errors.

While reducing above, FEG process maximizes the amount of work performed on-line as opposed to during planned outage while providing a link among Work-tagouts, the Maintenance Rule (MR), the Probabilistic Safety Analysis (PSA), and the Risk Monitor.

FEGs define dependencies between components and functions. Utilizing these dependencies will reduces effort necessary to assess the risk associated with the removal of components or a FEG from service.

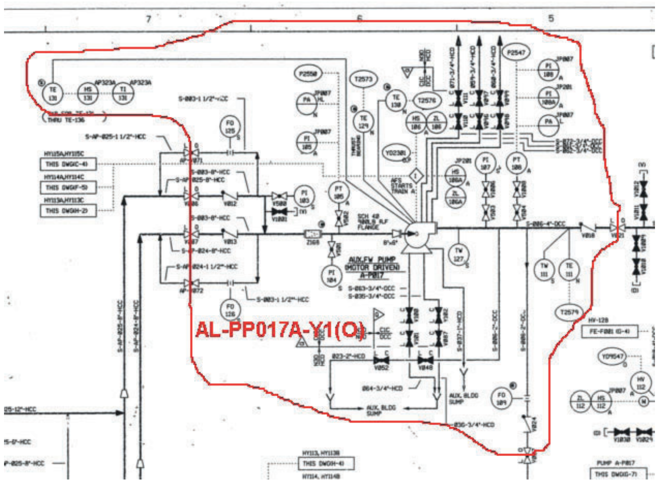


Figure 4 An example of FEG boundary drawing

Performance Management (PM) process is used to develop performance monitoring plans and to monitor the performance of plants, systems, component types, and programs according to the developed monitoring plans [15, 16]. PM process develops the utility Balanced Score Card (BSC), and establishes performance criteria and monitoring parameters for important system functions (System Monitoring Plan, SMP), component types (CMP) and programs (PMP). The BSC translates the vision and mission for the business into high level measures. Examples of performance measures included in BSC are regulatory agency's Reactor Oversight Program (ROP), WANO index, and plant/corporate Key Performance Index (KPI). BSC measures are used directly to develop lower level performance measures and to prioritize the consequences associated with health risks. Health risks include all threats to required performance level and opportunities to improve performance.

While SMPs are developed by system engineers, CMPs are developed by component engineers and PMPs are developed by program engineers. PM process ensures parameter monitoring that will identify symptoms of degradation and other types of health risks are being effectively, efficiently and consistently monitored. Most monitoring parameters are typically from condition monitoring parameters in the preventive maintenance programs which are developed in MO process.

Once any threat to plant health or gap in performance is identified, analysis of the causes is performed and required corrective actions are determined in CAP process.

SMPs are developed following completion of FID and FEG processes for a high priority group of systems. CMPs are developed prior to MO evaluation on each component types so that appropriate preventive maintenance program can be established in accordance with performance criteria in CMP.

Corrective Action Program (CAP) process is used to analyse root causes of any degraded performance identified in the PM process and to determine appropriate corrective actions.

For any type of component failure, it is directly entered into CAP process without PM monitoring. If the failed component has been classified as RTF, corrective maintenance tasks are performed without any causal analysis because RTF components are not in the scope of preventive maintenance program and only corrective maintenance strategy is applied on those. If the failed component has been classified as "critical" or "minor", consequences of the failure should be assessed whether any Maintenance Rule (MR) function is affected and tracked according to the MR program. Since the main purpose of preventive maintenance program is to identify performance degradation in advance of component failure and to conduct proactive actions, preventive maintenance program may present shortcomings for the failed component and the program is subject to a change as a result of the causal analysis.

The priority and depth of causal analysis is based on the functional importance of the affected component(s) determined in FID process. The more important component is affected, the more detailed root cause analyses are performed. Corrective actions (options for resolution) are determined as a result of the causal analysis. The options may include design changes, changes to preventive maintenance program, operational procedures, training program, or maintenance practices. All proposed options must be prioritized and the best option will be selected in Investment Management (IM) process. Action plans for the selected best option will be also developed subsequently.

Maintenance Optimization (MO) process is used to establish preventive maintenance program for each component type. The MO process is performed following completion of FEG, FID and PM process. FEG and FID processes need to be completed to ensure functional importance, component type, duty cycle, and service condition of each component have been defined in advance of MO process [17,18].

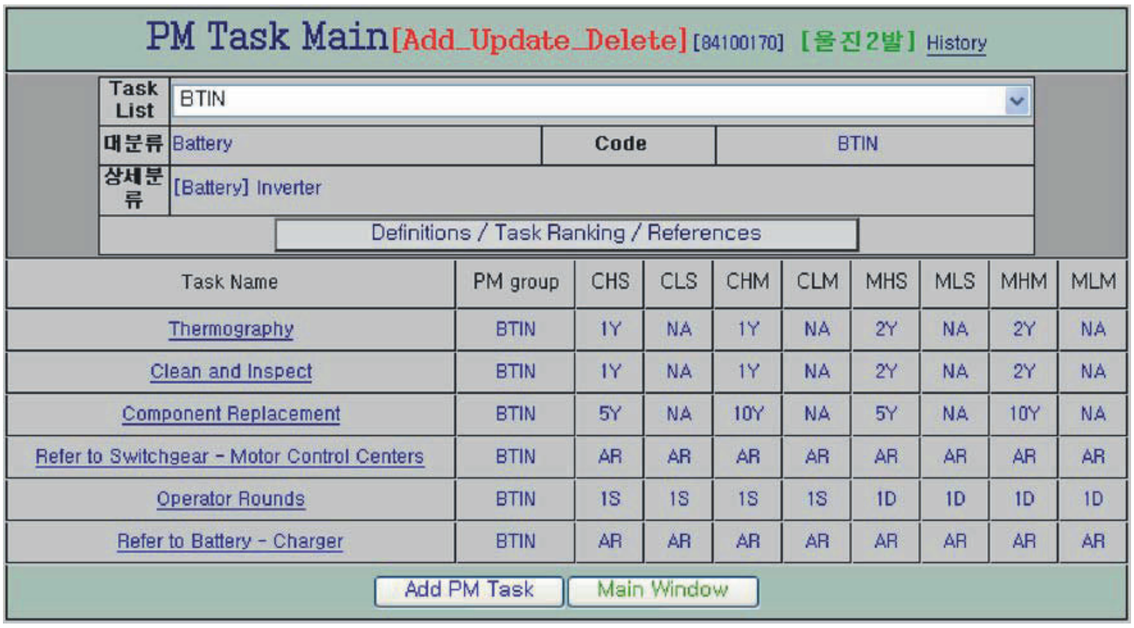
The key inputs to MO process are:

- FID results such as functional importance, component type, duty cycle, and service condition. This information can be downloaded from equipment recorder (equipment master in SAP PM Module) [19].
- Existing corporate/site preventive maintenance program (PM, PdM, condition monitoring)
- Vender recommended preventive maintenance tasks
- In-house/industry operating experiences
- Corporate/site specific performance history for the evaluated component type
- Performance criteria for the evaluated component type

PM templates are developed through MO evaluations for each component type and the format of PM template is the same as that of EPRI's PM template. The PM template includes technical basis that defines followings:

- Definitions for the component boundary, duty cycles, service conditions
- Failure locations, degradation mechanisms, and PM strategies
- Recommended PM tasks and their intervals for each MO asset group

MO evaluation for a specific component type is typically performed by an expert panel involving component engineers, system engineers, maintenance experts and external experts for the component type. MO asset group is the group of components which possess the same functional importance, duty cycle, service condition and PM basis. If any degraded condition is found in one component of an MO asset group, the other components in the same asset group are likely to be degraded and, therefore, need to be evaluated to determine whether the same degraded condition exist in them as well.


PM Task Main[Add_Update_Delete] [84100170] [울진2발] History

Task List: BTIN

대분류: Battery | Code: BTIN

상세분류: [Battery] Inverter

Definitions / Task Ranking / References

Task Name	PM group	CHS	CLS	CHM	CLM	MHS	MLS	MHM	MLM
Thermography	BTIN	1Y	NA	1Y	NA	2Y	NA	2Y	NA
Clean and Inspect	BTIN	1Y	NA	1Y	NA	2Y	NA	2Y	NA
Component Replacement	BTIN	5Y	NA	10Y	NA	5Y	NA	10Y	NA
Refer to Switchgear - Motor Control Centers	BTIN	AR	AR	AR	AR	AR	AR	AR	AR
Operator Rounds	BTIN	1S	1S	1S	1S	1D	1D	1D	1D
Refer to Battery - Charger	BTIN	AR	AR	AR	AR	AR	AR	AR	AR

Add PM Task | Main Window

Figure 5 An example of PM template

If MO evaluation results in significant increases in the scope of PM program, health risk associated with the results of the MO evaluation should be ranked and compared to other options in IM process prior to approval of the recommended PM program to select the best option to solve the performance gap initiating MO evaluation. Once approved, the results of MO evaluation are applied and implemented to preventive maintenance program for the MO asset group in SAP PM Module.

Work Management (WM) process is needed to perform and manage scheduled maintenance works [12]. This process would be developed and implemented after other processes are established. In the new process model, all works are scheduled in Long-Term-Planning process with considerations for nuclear risk (i.e. core damage frequency) which would be caused by performance of the work candidates. Operational risk monitor is used as a scheduling sub-tool to simulate and prospect risk profile. As a result of risk prospect, some work schedule might be arranged to lower risk peak under the limited line.

All works belonging to components of a specific FEG are grouped as a work package and treated together during work management steps such as work tag-out, work permission, post-maintenance test, and returning to normal position. Performing preventive maintenance works, technicians check whether the component is as degraded as expected in MO process or not and define as-found condition code according to component's degradation condition. The coded as-found condition information is fed back to MO process and used to optimise maintenance strategy. In case reported as-found condition is better than expected, we can extend task interval and/or eliminate the related task from PM template. In the other hand, we should consider to shorten task interval and/or to add effective task(s) into PM template if the condition is worse than expected.

4 LONG-TERM STRATEGIES

Investment Management (IM) process is used to evaluate and prioritize all of the recommended options that are proposed during CAP and MO processes to resolve the health risk [20,21]. Analytic Hierarchy Process (AHP) tool is being developed to support evaluation and prioritization of these options [22]. The action plans for the selected best option are developed and approved by Plant Health Committee. Once the action plans that need to be funded are approved, Appropriation Requests (ARs) are issued in the SAP IM module and passed onto budgeting process to ensure the funds and internal resources are available to support implementation of the action plans.

If the government's approved budget is not enough to support all ARs in the specific time frame, feasibility assessment for ARs may be re-performed and defer implementation of an approved but lower ranked ARs to another budgeting period. Since KHNP was invested mainly by Korean government, the amount of annual budget is controlled by the government.

All ARs, even deferred and unbudgeted, are a part of input to **Long-Term Planning (LTP)** process. Other parts of input to the LTP come from all work activities defined in previous processes such as FID, FEG, CAP and MO. Long-Term Planning (LTP) is a cycle schedule for the plant and developed based on FEGs using FEG characteristics and rules. The LTP is used to leverage the allocation of resources and risk across FEG work windows to reduce the number of times important functions are removed from service, and to optimize unavailability of these functions. The LTP becomes gradually elaborated through budgeting process. Although LTP development for the next cycle is typically completed in advance to ensure enough time to provide required resources, the minor work items may be added at the end of the current cycle of the plant.

Inventory Optimization (IO) process is used to define appropriate safety stock and track obsolete spare parts and materials [23]. The wider application of IM process can be called as Supply Chain Management (SCM) in modern business management.

The IO process ensures the right part or material is available at the right time at the right cost by maximizing the availability of parts and materials while minimizing the cost through innovative procurement and stocking strategies.

Equipment BOM information and functional importance are key inputs to IO process to define safety stock for the important materials used in critical components. The equipment BOM provides electronic linkages between components and materials required for the maintenance works. We can easily calculate how many components are subject to maintenance work and therefore determine required amount of materials, provided that equipment BOM has been accurately developed. The PM program established in MO process for the component type provides feedback to IO process on the frequency of certain material required for replacement or refurbishment tasks. Eventually the number of materials which should be in stock is estimated using equipment BOM and PM program. This number is then weighted with functional importance to define the safety stock for the material.

Other weight factors to define safety stock for the material are the lead time to purchase the material and obsolescence issues on the material. If the material can be purchased at the right time, the safety stock for the material can be decreased. Therefore all suppliers' capability to meet critical deadline for the delivery of importance materials must be demanded and tracked. If any equipment or material is already obsolete or seems to be obsolete soon, a search for alternate items should be initiated or decision should be made to replace the obsolete component with an equivalent model.

5 CONCLUSIONS

KHNP has been implementing a number of initial SAM processes such as FID, FEG, PM and MO. The FID, FEG, PM processes have been completed corporate wide while MO process has been implemented as a pilot project on a plant. KHNP also has plans to begin implementation of other processes such as CAP, IO, IM and LTP in this year.

The major advantages of SAM processes can be summarized as:

i. The limited resources are allocated on more valuable tasks for the importance assets through Graded Approach.

ii. The focus of maintenance is moved from the reactive to the proactive to reduce unexpected equipment failure causing high plant health risk and huge O&M cost.

iii. Analysis, prioritization and management of investment plans and preparation of contingency plans for the upcoming risks can be established in advance through Long-term Strategies.

KHNP has performed Business Process Reengineering (BPR) for several SAM processes as part of the ERP project in 2002. However, it took quite long time to implement SAM processes due to lack of organization's awareness for SAM concept. As time goes by, with continuous effort to educate and bring positive changes to the way KHNP works, most people began to perceive the advantages and acknowledge necessities of SAM processes. With SAM processes, many employees showed satisfaction at the value added works they performed. KHNP became greatly confident in the successful implementation of SAM processes which entail the advanced concept for managing physical asset based on graded approaches, proactive maintenance and long-term strategies.

6 REFERENCES

[1] W.B. Lee, et al., System Oriented Plant Maintenance, KHNP, 2004.

[2] W.B. Lee, Optimization of Maintenance Management for a Nuclear Power Plant Considering Its Safety and Economics, Bukyong University, 2005.

[3] W.B. Lee, et al., The Evaluation of Capital Investment for Improving the Performance of Aging Equipment in Nuclear Power Plant with AHP, KHNP, 2005.

[4] Anderson Consortium and KPMG Consulting Consortium, KHNP BPR Deliverables for MO, PM, FEG and IO processes, 2002.

[5] OnMarc Consulting Inc. & ACA Inc., *i*-ERIP process overview, 2005.

[6] INPO, Equipment Reliability Process Description, INPO AP-913, Rev. 1, 2001.

[7] NEI, The Standard Nuclear Performance Model - A Process Management Approach - Revision 4, 2003.

[8] EPRI, Critical Component Identification Process - Licensee Examples, TR-1007935, 2003.

[9] EPRI, Preventive Maintenance Basis Guideline, TR-106857 Vol. 1-38, 1999.

[10] EPRI, Preventive Maintenance Basis Database (PMBD) Version 5.1.1, Project 4109, 2004.

[11] NEI, Industry Guideline for Monitoring the Effectiveness of Maintenance at Nuclear Power Plants, NUMARC 93-01, 2000.

[12] INPO, Work Management Process Description, INPO AP-928 Rev. 1, 2003.

[13] EPRI, The Work Control Process Module in Support of a Living Maintenance Program, TR-108559, 1997.

[14] EPRI, Guidance for Developing and Implementing an On-Line Maintenance Strategy, TR-1009708, 2004.

[15] EPRI, Guideline for System Monitoring by System Engineers, TR-107668, 1997.

[16] INPO, Performance Objectives and Criteria, INPO 05-003, 2005.

[17] J. Moubray, RCM , Reliability-centered Maintenance, Industrial Press, 1997.

[18] J.D. Campbell & A.K. Jardine, Maintenance Excellence, Marcel Dekker Inc, 2001.

[19] B. Stengl & R. Ematinger, SAP R/3 Plant Maintenance - Making it works for your Business, SAP Press, 2001.

[20] NEI, Nuclear Asset Management Process Description and Guideline, NEI AP-940, Rev. 0, 2005.

[21] EPRI, Nuclear Plant Life Cycle Management Implementation Guideline, TR-106109, 1998.

[22] T.L. Saaty, Decision-making with AHP: Why is the Principal Eigenvector Necessary, European Journal of Operational Research, 2003.

[23] EPRI, Inventory Optimization in Support of the EPRI Work Control Process, TR-109648, 1998.

EXPERIENCE WITH EDUCATION IN MAINTENANCE AND RELIABILITY ENGINEERING BY OFF CAMPUS LEARNING

Ray Beebe

Senior Lecturer, Monash University School of Applied Sciences and Engineering

Abstract Physical plant assets are tending to remain in service for beyond their design lifetime. Arguably, a greater professional input is needed into the maintenance and reliability engineering of these assets, yet few universities offer such content in undergraduate engineering degrees. Accordingly, postgraduate programs have been developed. Those offered by off campus learning are more likely to meet the needs of engineers, many of whom work remote from the possibility of regular attendance on a campus. In the now 20 years since their inception, hundreds of engineers and other technical professionals have completed the Monash University programs. These have continued to evolve, and take appropriate advantage of advances in learning technology. This paper describes the development and lessons learnt during the Monash experience.

Keywords: education, maintenance engineering, reliability engineering, distance education, off-campus education, distance education, flexible learning, internet

1 INTRODUCTION

Many industrial assets are being required to not only remain in service beyond their intended design lifetime, but also to improve in performance. The focus of asset management has therefore moved beyond merely keeping plant functioning into the area of reliability improvement. A clear impact on the bottom line of a business must be shown. The term "maintenance" unfortunately implies a reactive role, but today it is increasingly seen as an investment in production capacity rather than just a cost to be borne.

Maintenance and Reliability Engineering is, therefore, clearly a professional engineering role offering great career challenge and excitement. Reliability Engineering is well developed in the military and in some technical areas such as electric power transmission. For the above reasons, these techniques are now relevant to industry in general. In Australia, the need for education of practising engineers in these areas led to programs taught by off campus learning.

Some universities around the world have recognised the need for such education, and have developed programs to suit. Some examples are: Hines (2005), Beebe and Houston (2001), Champaneri (1998), Knezevic (1997), Kerlin and Shannon(1997).

Off campus learning (sometimes called distance education) is the only practicable way of proving continuing professional development to those in this field, many at remote sites. It matches the needs of engineers wherever they live in the world. Also, research shows that:

"*Flexible learning is private learning based on course materials, providing access to teaching staff and two-way communication. Self-motivation is required, but the integration with the work scene leads to a higher capability outcome than if classroom-based" (Townsend & Cairns, 2003).*

2 ORIGINS OF THE MONASH MRE PROGRAM

The Graduate Diploma was developed in the mid-1980s by Len Bradshaw (who later left to publish MAINTENANCE Journal). The units of study were written by School staff, and differ in style in accord with academic initiative. Later, a Graduate Certificate was created for completion of the first four units.

In 1989, Dr Yousef Ibrahim developed a second graduate certificate: in Reliability Engineering. Funding supporting the development was from available Government and industry sources. It was written and is taught entirely by practising reliability engineers. The Master's degree was first offered in 1999, and designed to combine more reliability engineering material with the engineering maintenance management topics.

In 1999, an agreement was reached with the University of Tennessee at Knoxville, USA, to collaborate in offering these programs in North America. The UT Maintenance & Reliability Centre acts as the Monash agent, and arranges the Residential School in Knoxville, conducted by Monash staff.

3 THE PROGRAMS TODAY

The aim of the programs is to equip people with the knowledge, skills and attitudes to improve the life cycle performance of plant assets and thereby improve the performance of their business.

Study by off campus learning is usually half the study workload of on-campus students, but students can take a lesser load. For postgraduate continuing education, universities in Australia and the UK generally require one year of full-time study or equivalent for a *Graduate Diploma,* with a *Graduate Certificate* half of that. Requirements for Master's *degrees* vary. Two students completed the Master's degree in one year full-time, and this possibility may be attractive to others in a break between full-time work. The Monash programs in this field are:

- Graduate Certificate in Maintenance Management (GradCertMaintMgt)
- Graduate Certificate in Reliability Engineering (GradCertRelEng)
- Graduate Diploma in Engineering Maintenance Management (GradDipEngMaintMgt)
- Master of Maintenance and Reliability Engineering (MMaintRelEng)

Direct admission into the Master's is available to engineers with an Honours degree in engineering, or with a pass degree and honours equivalent experience. Admission into the Graduate Diploma is available to lower grade graduates in engineering and graduates in other professions. Others may articulate following completion of the Graduate Diploma with high grades.

There are many non-degreed engineers and others working in maintenance and reliability, often with significant responsibility for staff management or planning of maintenance. Anybody enrolling direct with Monash can study units on a "not-for-degree" basis. On completion of two units with an average grade of 65%, admission into either of the Graduate Certificates is allowed with full credit. Completion of either Graduate Certificate with 65% grades entitles articulation into the Graduate Diploma and later the Master's degree. Many such people are of high calibre and some have completed Master level.

Assessment varies with the units, with all having progressive assessment via assignments submitted through the year. Most have an examination as well, are arranged anywhere in the world. Some units have assignments done in groups at the Residential School (see later).

Students usually study four units per year, but some choose less. An unusual feature compared with other programs and other universities is that the teaching period for all units is the full year, from late February to October, rather than the more usual semester basis. This gives added flexibility to manage any workplace, personal or family circumstances.

Some hundreds of people have completed the programs, and their feedback is overwhelmingly positive. In 2005, most came from every State in Australia, but also from 13 other countries. In 2006, the programs have students in every inhabited continent. Wherever they live, students interact with teaching staff in exactly the same way as those in Australia.

4 STUDY MATERIALS

Based on feedback from students, the study materials continue to be produced in print based format, updated each year and rewritten regularly as appropriate.

There are generally three parts, with some units also having a CD of images, short videos or other resources:

- Unit Guide, which contains information about unit administration and gives the assignments and other assessment detail (many assignments relate to the student's workplace);
- Unit Book, which contains the Study Guides. These are detailed guidance to the subject matter, referring to the textbook and Reader/s where appropriate. The style of these varies with the topic and text;
- Reader, consisting of selected relevant technical papers and articles. Although items are reviewed annually, some items remain that are regarded as classic in the field.

Most units require a textbook, which is usually bought by the students. An exception to the above is the Maintenance and Reliability Engineering Project, which is workplace-related.

5 RESIDENTIAL ATTENDANCE

Study in the maintenance management units requires participation at one Residential School of 3.5 days in each year at the Gippsland Campus in Churchill. The School for the first four units is held in March-April. A feature of the September School is the presentation by students of formal papers on their project. With upwards of 20 such papers, much is learnt by all in what is effectively a quality mini-conference.

The Schools have activities from day into evening. Contents include lectures, specialist guest speakers, student presentations, group activities, and usually a relevant site visit. The Schools are regarded as essential by the students, with networking being a highly valued component. A single School each year is run in Knoxville, combining the above features.

6 TEACHING

Unit Advisers comprise full-time academics, who also have other teaching, management and research commitments, and sessional staff, who are practising engineers. Some have international reputations in their field. The author has been Co-ordinator since 1996, joining 4 years earlier following 28 years in the power generation industry.

Students maintain contact with the Unit Advisers mostly via e-mail, and less so by telephone. Assignments are submitted by e-mail, and are responded to with feedback by the Unit Adviser after grading.

The program is offers a graded coverage of subject content, designed to provide students with knowledge, skills and attitudes in each subject area. Students are expected to read widely, and present responses to assignments which indicate their understanding and appreciation of the application of the material. Sometimes there are no clear right or wrong answers!

Continual additions are made to the Monash Library collection in the MRE field, and students are encouraged to use its mailing out service. Use of other libraries is available where co-operative arrangements have been agreed with several other institutions. Electronic resources are a rapidly growing source, with electronic journals and databases enabling access anywhere in the world at a time that suits the students. Many textbooks are available in electronic versions, where each "copy" held can be accessed on line by one person at a time. Some students have access to texts and other resources held by their companies.

7 THE MOVE TO WEB-ENHANCED TEACHING

As with many other aspects of life, the computer and internet also provide great possibilities for teaching, with the aim of improving quality (Benson, 2003). Our associates at the University of Tennessee developed six units of study in this field and trialled them successfully (Hines and Shannon, 2000). A complete master's degree in a closely associated field is now proposed (Hines, 2005). Teaching takes place synchronously, i.e. with the students all logged on at the same time. Students and teacher can communicate with each other online. The teacher's presentation is recorded for later access by students.

Monash University has adopted WebCT VISTA for all units, following initial use of WebCT CE. This is called MUSO, for "Monash University Studies Online" and is the generic term used in the rest of this paper. In the interests of extending off campus delivery, development funding was provided for this program. Each Unit Adviser as the content specialist was assigned an Educational Designer, qualified and experienced in teaching. The Unit Advisers and Educational Designers worked together to learn MUSO and develop the detail of the units.

The three broad pedagogical approaches identified for development of MUSO templates rely jointly on adult learning principles and resource-based learning, and are (Bruhn, 2004):

- Basic collaboration – where teaching is to complemented with some basic collaborative tools (e.g. discussion groups);

- Case-based or Problem-based – where these philosophies are used in a unit to give selective release of cases and flexibility in organising learning environments to support students working in small groups, and the style adopted here:

- Mixed mode – for on-line interaction, provision of content offered on campus or via off campus print formats, and uses various tools to support learning: the approach taken with these units.

8 FORMULATION OF MINIMUM STANDARDS FOR WEB-ENHANCED DELIVERY

In devising a minimum standard that should apply to the units when adopting a mixed mode approach to the delivery of units online the following criteria (Bruhn, 2004) were developed for the WebCT site(s) to:

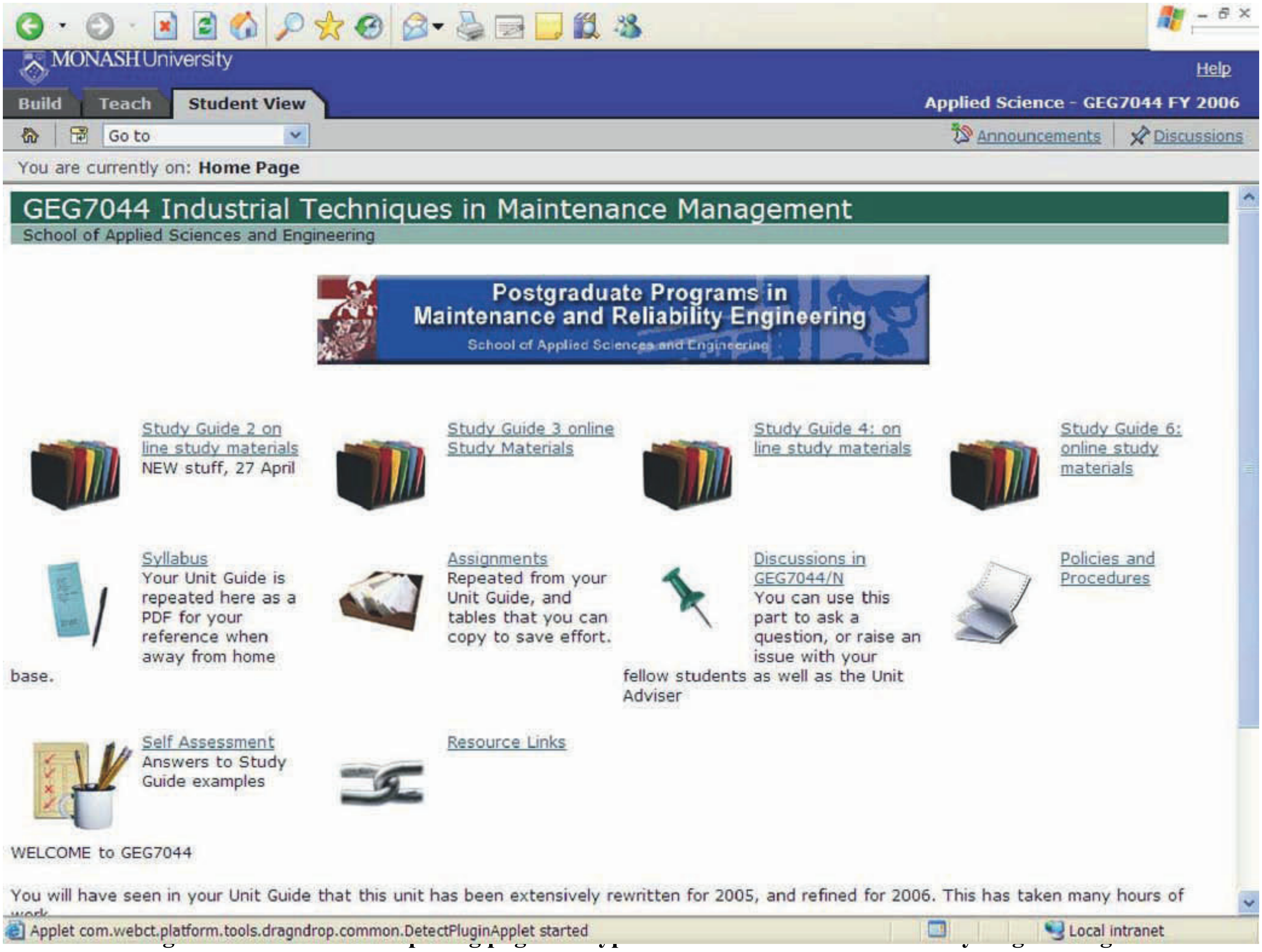
- Easy to navigate and be strong on 'visual signposts', that is, graphics and/or text on the web pages should be self-explanatory as to what will be presented if the student clicks on that icon.
- Have clearly expressed objectives or learning outcomes to define content, assessment tasks (both formative and summative) and online learning activities.
- Provide content in a form appropriate for online delivery, or give directions/guidance as to where it is located.
- Complement/enhance any print materials and not just duplicate them.
- Provide opportunities for 'active learning' through the careful integration of interactive learning tasks (which may include interactive multimedia), that is, the students should engage in activities which require them to do something (e.g. role play within a game, manipulate models by changing a range of variables, visit a web site and do a critical appraisal of its features).
- Provide opportunities for students to assess their progress in acquiring knowledge and skills through the use of quizzes and other online self-assessment activities (formative assessment) and as preparation for summative assessment tasks.
- Provide an efficient means of linking to other online resources both within and external to the University and have the capacity for the academic to provide up-to-date information that becomes available during the semester.
- Provide the means by which students can communicate with the academic or fellow students, either informally or through structured discussion groups.
- Provide the means by which students can access important information about the unit and critical dates during the semester [e.g. unit structure, prescribed texts, assignment details (including assignment and examination dates)].
- Be flexible in design to allow for changes to be incorporated as required during the semester based on the identified needs of particular cohorts of students and/or variations or additions to the content.

9 THE DEVELOPMENT PROCESS

It was decided early that internet use would be non-synchronous, as the students are spread across many time zones. It was also firmly established that the development was fundamentally being made to improve the quality of teaching and learning, not to save money, nor to just appear to be progressive. Active learning was the aim, through careful integration of learning tasks, so that a student would be required to do something, rather than be a passive reader. In some cases, the existing materials included these features, and enabled ready adaptation into a web-enhanced approach. Multi-media expertise was utilised where appropriate.

It was also decided that the study materials would remain in print-based format, supplemented with CDs for sound, images, and some documentation. All students are required to have internet access, although it is still not expected yet that all students have broadband access. Merely transferring the study materials away from provision in print to having them on-line, such that the student needs to download and probably print them, is not the best use of the internet. High quality images open quickly from a CD, but can take some time to download if on line.

The opening page of one unit is shown in Figure 1. The material content can be built up and readily modified by the Unit Adviser, who has "Designer" access.



Content pages vary with unit and teaching style, but may comprise:

- Syllabus: an outline of the aims of the unit, its assessment and administrative details.

- On-line study materials: a section corresponding to appropriate Study Guides (up to 8, depending on the unit).

- Self Assessment Questions: each Study Guide contains Learning Activities, with responses given at the end of each print section. Self Assessment Questions are available for each Study Guide, and are answered on line. After the student enters a response, the "official" answer is shown. These can be graded, but it was decided not to do this, but to use them as opportunities for students to assess their progress in acquiring knowledge and skills.

- Resources: a display of hot links to selected web sites with information relevant to the unit material. These can be chosen to open in a new window.

- Communication: consisting of two components: Discussion groups for interaction between students

Assignments: a summary or repeat of the requirements given in print.

10 EXPERIENCE

The extent of site development and particular features varied between units, as it was not desired to have the teaching pattern standardised, but to allow for individual teaching and presentation styles, as with all university teaching.

The varying number of students was also a factor, such that experiences differed between units. It took some time for both teachers and students to get familiar with MUSO, and several had some difficulty utilising all the required features. MUSO provides on line facility for drafting and eventually submitting assignments. A back-up arrangement for submission of assignments had been set up, and was in fact used by most. The intent was intended that Unit Advisers read the assignments, added comments and assessment grade on screen without printing them. Most managed to obtain this skill. However, submission by email direct to the Unit Adviser has proven to be simpler for all concerned and is now our standard method.

Towards the end of the second year of offering the units this way, formal evaluation was performed by a Senior Education Designer who had not been involved. A questionnaire was used to evaluate student experience in each of five units. For one unit, a two-hour focus group was also held. Sixteen of the 30 students in this unit responded, with some points that emerged as follows:

- Time is needed to obtain sufficient familiarity with MUSO and the unit site. Several teething troubles were encountered, and most overcome as the year progressed.

- The on-line resources were inconvenient to access compared with the print materials, and some saw as merely being a repeat of the paper materials and not adding value to them.

- Not all could access the internet when away from home when they wished to do some study.

- The volume of reading concerned some, and it is possible that this may be less of a problem for those who had previous tertiary study. Some felt that the site should include add-on information to the print materials. Student-selected resource links were suggested by some.

- Discussion groups did not occur because students did not raise issues. These need to be set up by the Unit Adviser. An early one could be for students to post introductory information about themselves (as most had never met). Although there are several public discussion boards in the maintenance and reliability engineering field, these are apparently not used as widely as it appears. As many students travel away from internet access in the course of their employment, structured discussions may be difficult to introduce.

- The self assessment questions were used by some for revision prior to the examination (which was conducted in the normal fashion). Some felt that if they had been graded and thus included in the overall assessment for the unit, they would have been given earlier attention. General experience seems to be that is that items will not be attempted unless they are required to be completed, whether assessed or not.

- Many favoured an introductory video to accelerate the learning curve for the start of the unit.

- Overall, the students rated the unit as “very good”, and much was learnt to improve it for future years.

Later, all of the units were rated against the formulation criteria. This unit fully met the criterion in 7 of the 10, and partially in the others.

11 WHO TAKES UP MRE STUDIES?

In 2005, a questionnaire was developed in conjunction with Dr Melinda Hodkiewicz of UWA and sent to a range of current and recent students. 87 responses were received and analysed. The average working life of respondents was 16.7 years

The **discipline** of engineering etc. most strongly represented was Mechanical (66%), followed by Electrical (16%), Electronic/mechatronic (7%), Management (3%), Other (6%) and Chemical (2%). 25% of respondents **described their jobs** as "reliability engineers", 47% are in "maintenance" (Engineer, Planner, Supervisor, Technician) and 23% are "managers". The two occupational terms are not used consistently. Hodkiewicz et al (2004) built on O'Malley et al (2003) and suggested that the distinction should be:

- Reliability engineers are not governed by day-to-day operations, but have a strategic focus. They develop programs that provide the framework for maintenance.

- Maintenance engineers are responsible for immediate equipment functionality and implement strategies developed in co-operation with reliability engineers.

Table 1 shows the **industry** where students in this sample work by percentage:

Mining/mineral processing/refining	37	Petrochemical	5
Defence (military & civilian)	17	Government	2
Manufacturing	15	Warehouse/logistics	2
Utility (gas, water, power)	10	Timber/pulp & paper	2
Service Provider/Consultant	8	Transport (air, road, rail)	2

12 WHAT DO STUDENTS CONSIDER SHOULD BE INCLUDED IN MRE COURSES?

The topics considered were in two categories. Each topic had some students who considered that they were competent already (remember that these are not raw engineers). The survey results are shown in Figure 2 (technical topics) and Figure 3 (management topics). Note that some of these are not included in traditional bachelor degree programs.

The Top 3 technical were: Reliability and Risk Based Inspection methods
Failure modes and effects analysis
Failure mechanisms & Root cause analysis

And the Top 3 management: Life cycle costing and NPV methods
Risk assessment
Budgeting and capital expenditure

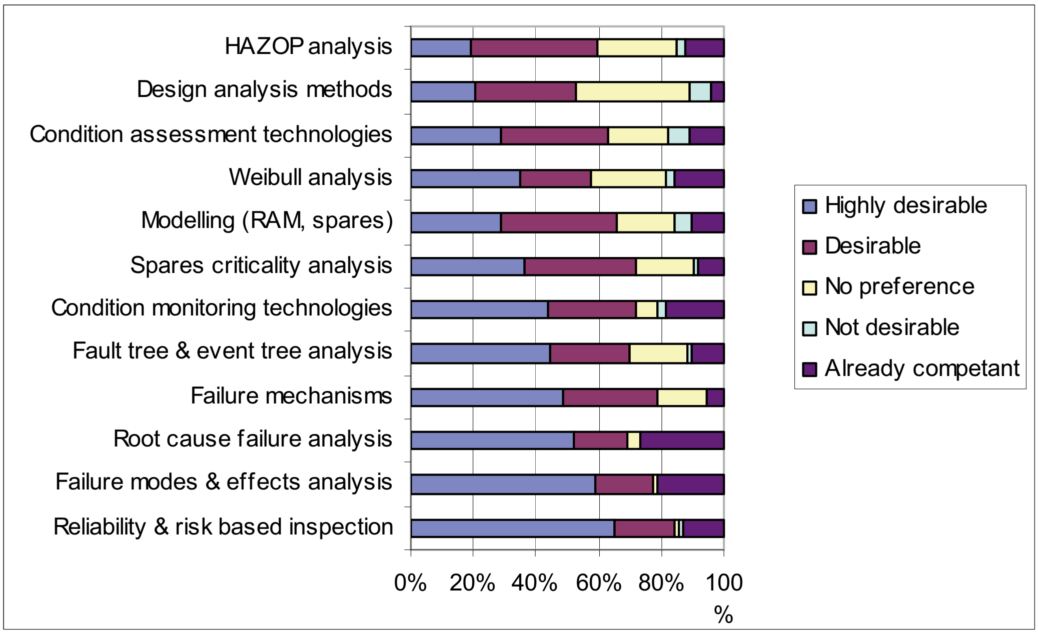


Figure 2: Value of technical topics in MRE study according to student survey (N=87)

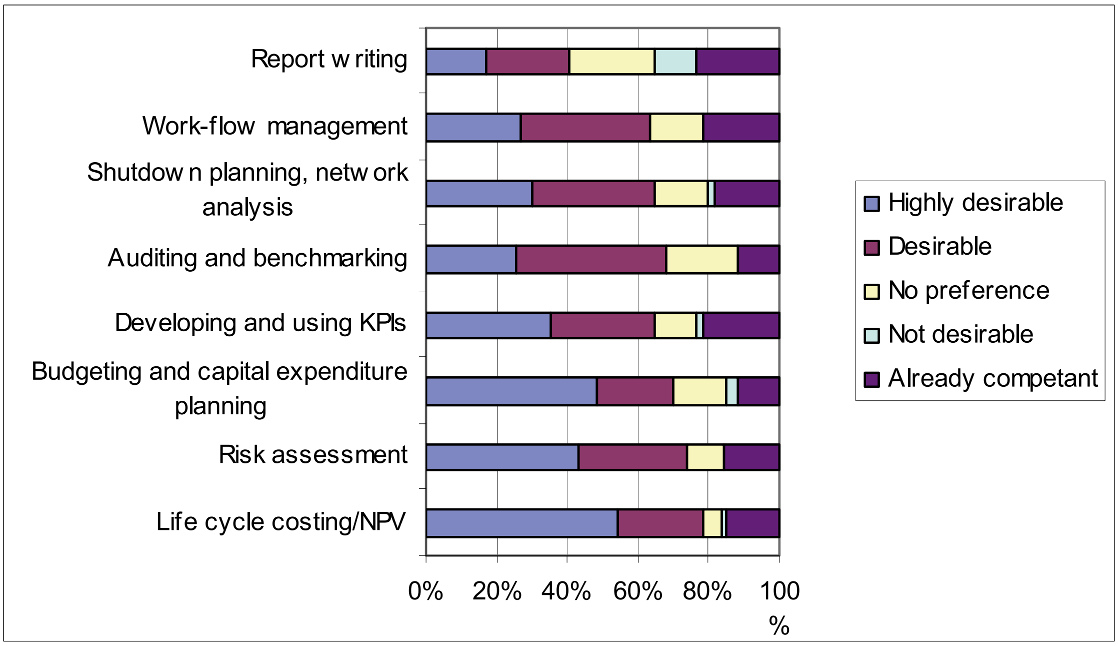


Figure 3: Value of technical topics in MRE study according to student survey (N=87)

13 WHAT OUTCOMES HAVE GRADUATES EXPERIENCED OR EXPECT TO EXPERIENCE?

Students have not been asked this until now, but reports have been received over the years of promotion and new jobs where the studies have been a major factor. Figure 4 shows the survey results. Worthwhile outcomes were apparent across the board, although the data could be interpreted as showing an individual highly satisfied in all areas, or in only some!

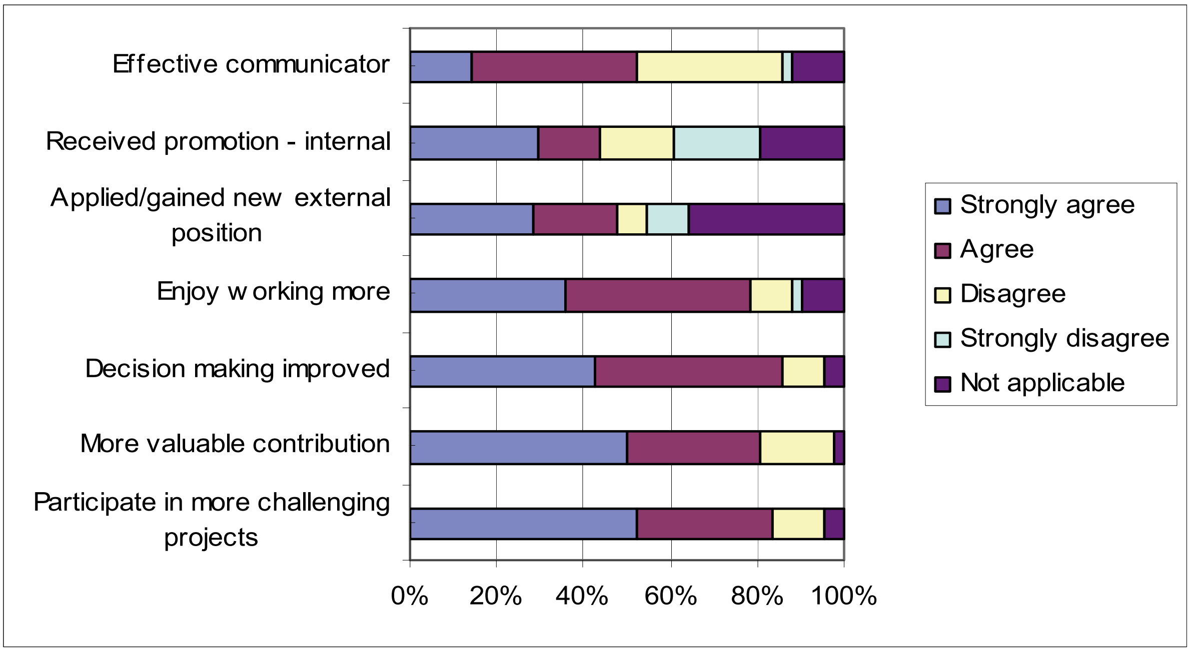


Figure 4: Outcomes experienced or expected from MRE study according to student survey (N=87)

14 CONCLUSION

Experience with web-enhanced delivery of off campus learning in maintenance and reliability engineering shows that improved learning is possible, but significant overall benefits should not be expected initially when compared with paper-based print materials alone. This should improve as higher performance internet access becomes more widely available.

In developing a unit of study for web-enhanced delivery, the academic must try to use opportunities possible on-line that cannot be provided easily in other ways. These must be integrated fully with all the other components of the study materials in a unit in line with the overall objectives of that unit: the only reason for having web-enhanced learning is to do a better job.

The survey gave some insights into what students consider about the value of topics in MRE programs, as well as the value of postgraduate study itself. It is overwhelmingly thought to be worthwhile, and gives the basis for further research into topics that should be included or emphasised. Some points:

- Reliability is a job people “move” into, not an early career choice.
- Those in the reliability field have a wide variety of educational backgrounds and experience.
- Considering the population of engineers, relatively few have undertaken PG reliability education.
- Many reliability professionals undertake targeted education in the form of university-coordinated postgraduate studies, vendor- supplied or institution organised training.

15 REFERENCES

[1] Beebe, R and Houston, D (2001) *Education in maintenance and reliability engineering at a distance* Proceedings of ICOMS2001 (International Conference of Maintenance Societies), Melbourne, pp 038.1 - 038.9

[2] Benson, R. (2003). *Improving the quality of distance education through online learning* in G. Davies & E.Stacey (Eds.), *Quality education @ a distance* (pp. 169-178). Boston: Kluwer.

[3] Champaneri, R (1998) *Glasgow Caledonian University's Contribution to Maintenance Management Programmes* Proceedings of 11th COMADEM, Launceston, Australia, pp 141 -147

[4] Hines, W and Shannon, T (2000) *The future of maintenance and reliability engineering education* Proceedings of 13th COMADEM, Houston, pp 189-193

[5] Hines, W (2005) *Master's degree in reliability and maintainability engineering* Maintrain2005 Annual Conference Plant Engineering and Maintenance Association Canada (PEMAC)

[6] Hodkiewicz , M, Sikorska, J and Simpson, P (2004) *Creating the next generation of reliability engineers* Proceedings of ICOMS2004 (International Conference of Maintenance Societies), Sydney, pp 039.1 – 039.12

[7] Kerlin, T and Shannon, T (1997) *A University's response to the industrial maintenance and reliability challenge* Proceedings, MARCON pp 74.01 - 74.05

[8] Knezevic, J (1997) *Industry driven postgraduate maintenance education: MIRCE approach* Journal of Quality in Maintenance Engineering, Vol 3 No 1 pp302 -308

[9] O'Malley, A and Karyagina, M (2003) *The role of a reliability engineer* Proceedings of ICOMS2003 (International Conference of Maintenance Societies), Perth, pp 039.1 – 039.12

[10] Townsend, P and Cairns, L (2003), *Developing the Global Manager using a Capability Framework*, International Journal of Management Learning, Vol.34(3), p.333-347

Website: www.gippsland.monash.edu/science/mre

Acknowledgements

The close working of teaching, designing, and multimedia colleagues is appreciated as is the feedback received from the students, all of which will help improve the teaching and learning experience.

MONITORING CENTRAL GLAND LEAKAGE ON COMBINED HP-IP CASING STEAM TURBINES

Ray Beebe

School of Applied Sciences and Engineering, Monash University

Ray.Beebe@eng.monash.edu.au

Abstract: Steam turbine generators, the mainstay of electricity production, are increasingly required to run for well beyond their intended lifetime. As opening up machines for inspection is expensive, owners need to consider all relevant information in making the decision. Deposits on blades and erosion of internal clearances reduce machine efficiency and output, and can be detected and monitored using condition monitoring by performance analysis. Data obtained before and after overhaul also reveals whether any restorative work achieved the expected improvements in performance. Many large machines, including an increasing number in Australia, are of the compact combined High Pressure–Intermediate Pressure opposed-flow casing design. Increasing internal steam leakage from entry into the high pressure section to the intermediate pressure entry section has a large effect on output and efficiency. This flow through the central gland packing cannot be measured directly. Using a case study, this paper describes a method of estimating it by test that is sufficient for routine condition monitoring. The test explained observed poor performance that had led to extensive modifications to the boiler.

Keywords: steam turbines, diagnosis, condition monitoring

1 INTRODUCTION

Steam turbine generators, the mainstay of electricity production, are increasingly required to run for well beyond their intended lifetime. As opening up machines for inspection is expensive, owners need to consider all relevant information in making the decision. Condition monitoring by performance analysis is one important tool to help in such major maintenance decisions, and can point to the presence of deposits on blades and erosion of internal clearances that reduce machine efficiency and output. (McCloskey et al, 1999). Data obtained before and after overhaul also reveals whether any restorative work achieved the expected improvements in performance.

Basic condition parameters are the Valves-Wide-Open Output and Enthalpy Drop Efficiency (ASME, 1985 and Beebe, 2003). Tests are required with special calibrated instruments, but recent experience shows that useful results can be obtained on units with DCS (Digital Control System) (Beebe, 2005).

Most large steam turbine generators have separate High Pressure and Intermediate Pressure casings, followed by one or more Low Pressure casings, virtually always in opposed flow (also called double-flow: steam flows in opposing directions from admission). The greater the number of casings, the more expensive the machine and the longer the building to house it. Advances in design and materials of Low Pressure blading have led to longer last stage blades, such that a single LP casing is now possible in sizes of up to 500MW.

Another design layout that results in a more compact machine is when HP and IP turbine sections are combined in one casing in double-flow style. This arrangement is used in many sets in Australia from 350MW to 500MW capacity and elsewhere up to larger capacities. These sections have a central gland between them, sometimes known as the N2 packing (Figure I).

Even in good condition, due to the difference in steam pressures, steam leaks from the HP turbine through this gland into the IP inlet area. This leakage must be known, or at least estimated, to obtain the heat rate (i.e. inverse of the thermal efficiency) of the turbine at the acceptance test. It also reduces the effective enthalpy of the IP section steam flow, so that the enthalpy drop efficiency on test can give a misleading indication of IP blading condition (Cotton, 1993). This is important in condition monitoring of thee type machines.

This paper describes experience with a design of large steam turbine that had a poorer heat rate on their acceptance tests than expected. In early operation, difficulties were experienced in the boiler that required modifications. Special tests run later on the turbines revealed the internal leakage to be a major root cause of the boiler problems.

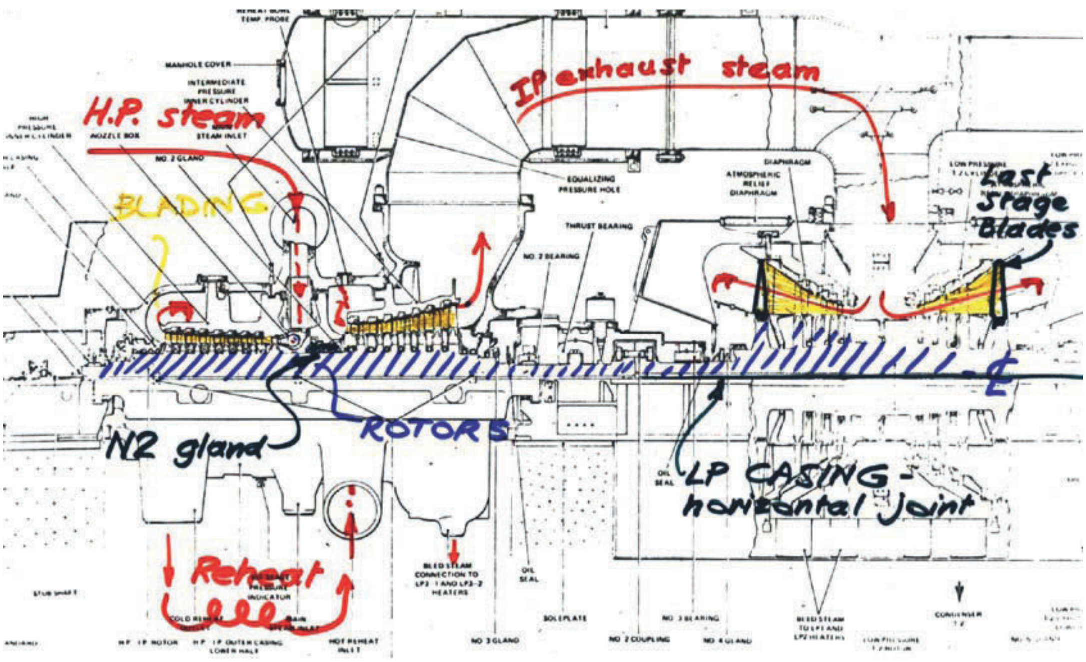


Figure I: cross-section of a typical HP-IP turbine.

2 IMPACT OF N2 PACKING LEAKAGE

This leakage is usually taken as about 1% of main steam inlet flow, although has been found to be reduced to 0.5% with a casing retrofit (Hogg & Stephen, 2005). This leakage, and any increase, means that some of the steam initially supplied to the turbine does not pass through the HP blading, then through the reheater, and back to the IP turbine, but instead directly to the IP blading. As the leakage flow effectively circulates in a non-reheat cycle of lower efficiency, N2 leakage can have a dramatic effect on turbine heat rate/thermal efficiency and output. Typically, an increase in N2 leakage of 1% of initial steam supplied to the turbine results in an output reduction of about 0.3% and efficiency reduction of about 0.16% (Koch, et al 1988).

Further leakages may occur from other places, according to the design of the machine, such as through eroded or distorted horizontal joints, and from around the main inlet pipe piston-ring type seals are possible.

As mentioned, the enthalpy drop efficiency of turbine sections is a prime parameter (ASME, 1985 and Beebe, 2003). As the N2 leakage flow enters after the steam temperature into the IP casing is measured, and is at the enthalpy corresponding to the value from the HP turbine expansion at the first stage pressure, it is lower than the enthalpy of the steam supplied to the IP turbine from the reheater. Hence the steam entering the IP blading at the IP bowl in total does not have the enthalpy of the main inlet mixture. The result is to give an incorrect high value of IP enthalpy drop efficiency. Therefore, in routine testing of such turbines, if an increase in IP enthalpy drop efficiency is found, then increased N2 gland leakage should be suspected.

Most turbines have interleaved labyrinth type seals, but a later brush seal technology has been developed by one manufacturer (Hogg and Stephen, 2003). The HP-IP rotors are usually flexible, such that they pass through a bending critical speed during starting up and coastdown operation up to or down from normal service speed. As this gland is near the centre of the rotor span, the consequent rotor deflection can cause rubbing, and thereby increase the clearances. Another packing type available new or at retrofit is retractable to prevent rubbing. They close in to running clearances when the turbine is at service speed.

3 HOW TO FIND THE LEAKAGE FLOW

Methods of finding the leakage steam flow are given in Cotton (1993) and industry papers not published for general access. Experience such as reported by Koch, et al (1988) and Staggers and Priestley (1990) has led to an EPRI research project (Haynes et al, 1995).

It cannot be measured directly without installing a bypass pipe around the emergency blowdown line, and including a flow measurement device. (The blowdown line and valve are fitted on most designs to release steam entrapped in the casings when the machine is tripped).

Another method requires IP enthalpy drop efficiency tests to be run before and after opening the packing blowdown valve. Operating this OPEN-or-SHUT valve carries some risk of damage to packing teeth. Use of this method was reported by Hogg and Stephen (2005) where the risk was evaluated as acceptable. A bypass line was fitted to obtain finer control, but later left closed and full blowdown flow enabled.

Other ways of estimating it to sufficient accuracy for routine condition monitoring are given in Cotton (1993) and Booth (1984).

One method described there requires neither modifications nor risks to the plant. Tests are run at steady conditions with some differences between main steam and reheat steam temperatures arranged. This is readily done at Valves Wide Open by holding the hot reheat steam temperature constant while reducing the main steam inlet temperature, and vice versa. Three tests are suggested. At each test, the IP efficiency is calculated, and plotted against two assumed leakage flows. The intercept enables a reasonable estimate of the leakage flow (see Fig II).

The changes in steam inlet temperatures are also likely to alter gland clearance (Koch, et al 1988). A critical assessment of the accuracy is given in Haynes, et al (1995), but it is sufficiently accurate for an economic analysis to justify the required replacement parts to be obtained and the job scheduled as part of a planned major outage.

Recent experience (Beebe, 2005) shows that on plants with DCS, and as useful information can be gained using the permanent instrumentation system for overall condition monitoring, plant owners could well investigate its use to monitor N2 leakage and save on test cost.

Steps in the process are:

1. Run a Valves Wide Open test at as close to design HP and IP inlet steam conditions as can be achieved. Ensure that steam temperature and pressure are measured at inlet to the reheat intercept valves, and at the IP turbine exhaust (where experience may confirm that more truly representative values are obtained at the inlet to the LP turbine: Cotton, 1993).
2. Using the measured IP inlet steam conditions, calculate IP Turbine Enthalpy Drop Efficiency (steam properties tables are available on the internet calculate). Plot this on the vertical axis of a graph of efficiency vs N2 leakage, for zero leakage.
3. Assume a value of N2 leakage as a percentage of IP bowl flow, and calculate by heat balance the enthalpy of the steam to the blades:

Heat flow into the IP bowl = N2 packing leakage + Heat flow at IP inlet valves = Heat flow to blades from IP bowl:

$$m_1 \times h_1 + m_2 \times h_2 = m_3 \times h_3$$

where:

- m_1, h_1 = mass flow and enthalpy of steam leaking into IP bowl through the N2 gland, taken from the HP turbine expansion line at first stage pressure.
- m_2, h_2 = mass flow and enthalpy of hot reheat steam entering the IP turbine at the valves.
- m_3, h_3 = mass flow and enthalpy of steam through the IP blading (i.e. that of the mixture of inlet flow and leakage)

It is assumed that the flow m_3 through the IP blades is constant (i.e. the greater the N2 leakage, the lesser the IP inlet flow m_2 by the same amount). For a given leakage, the flow m_3 is % leakage × m_1

By manipulation, for an assumed leakage flow of 10% of the flow m_3 :

$$h_3 = h_1 + 9\,h_2 \qquad\qquad 10$$

4. Use this enthalpy value to again calculate the IP enthalpy drop efficiency, and add the result to the plot at 10%. Make further similar calculations if desired for other % leakage values. Draw a line joining the points, as shown diagrammatically for Test 1 in Figure II.

5. Run a second test at constant IP inlet temperature, but with a lower main inlet steam temperature, within the range allowed by the turbine manufacturer's instructions. Cotton (1993) gives 40 C° below design as practical. Repeat the calculations given above. (Test 2)

6. If desired, run a further test with design main inlet steam temperature and a lowered IP inlet temperature. (Test 3)

The intersection of the lines gives a reasonable estimate of the actual leakage flow (Figure II). A close estimate is also given of the true IP Enthalpy Drop Efficiency from this intersection. A cycle modelling program can be used to calculate the impact on output and thermal efficiency of any increased leakage, and used to justify restorative action or modifications.

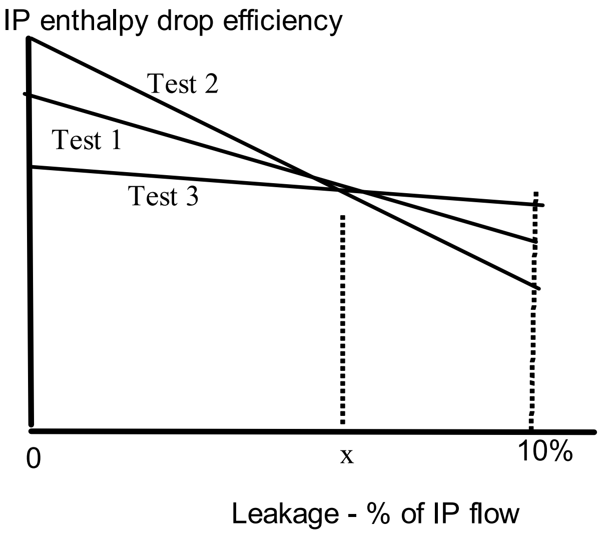


Figure II: Example plot of results for N2 gland leakage: x is estimated leakage flow

4 THE CASE STUDY

When a turbine generator of the combined HP-IP casing design was tested for acceptance, it was found to have a thermal efficiency well below that expected. The test procedures, particularly the primary flow measurement, were closely examined, but no discrepancy could be found and no reason for the shortfall was agreed. In operation of this unit, there was also difficulty in holding hot reheat temperature down to the design values. To ensure operation at design temperature, an extra desuperheating temperature control spray station was designed and installed in the boiler and 20% of the tubing in the first section reheater was removed.

Information on the N2 packing leakage tests described in this paper became available some years later, and when some routine VWO tests were arranged, it was decided to include those tests. Table I summarises results for tests of type 1 and 2.

Table I	Initial test at design conditions	Test with HP inlet temperature lowered
Main inlet steam pressure and temperature	13422.7 kPa Abs, 532.5 °C	13274.6 kPa Abs, 497.51 °C
IP inlet steam pressure and temperature	4058.8 kPa Abs, 537.5 °C	4046.6 kPa Abs, 538.1 °C
IP exhaust steam pressure and temperature	942.1 kPa Abs, 329.6 °C	951.3 kPa Abs, 329.14 °C
Enthalpy of main steam	3418.9 kJ/kg	3326.47 kJ/kg
Enthalpy of first stage steam h_1	3386.4 kJ/kg	3296.4 kJ/kg
Enthalpy of IP inlet steam h_2	3529.6 kJ/kg	3531.0 kJ/kg
Enthalpy in IP bowl if N2 leakage is 10% of IP flow h_3	3515.3 kJ/kg	3507.5kJ/kg
IP enthalpy drop efficiency with 0% N2 leakage	90.81%	91.98%
IP enthalpy drop efficiency if N2 leakage is 10% of IP flow	88.44%	88.07%

The estimated N2 gland leakage flow was found to be between 15 and 22 kg/s, or 4 to 5 times the design value. The design value had of course been used in calculating the steam flows to and from the turbine at the acceptance thermal efficiency tests. This increased leakage flow, amounting to a reduction of between 5.5% and 8% in the flow from HP turbine exhaust to the reheater, was consistent with the reduced thermal efficiency and reheater problems.

Inspection at the first overhaul some years later confirmed that the N2 gland clearances were greater than design, and had presumably been so since the acceptance tests and probably since new. Later, retractable packings were installed, but the repeat check test showed that the leakage was still present.

5 CONCLUSION

For steam turbine generators of the HP-IP casing type, N2 packing leakage tests should be part of the routine condition monitoring by performance analysis program and conducted at acceptance and with every Valves Wide Open test. If significantly increased N2 packing leakage is detected, the accuracy is sufficient to make an economic case to obtain replacement gland parts for the next planned major outage, or of conversion of the gland packing to the retractable type.

6 REFERENCES

[1] ASME (1985) Simplified procedures for routine performance tests of steam turbines ANSI PTC 6S report 1974, reaffirmed 1985.

[2] Beebe, RS (2003) *Condition monitoring of steam turbines by performance analysis* Journal of Quality in Maintenance Engineering Vol 9 No 2 April Page: 102 – 112

[3] Beebe, RS (2005) *Steam turbine performance condition monitoring using plant instrumentation: case study* Proceedings of International Conference of Maintenance Societies ICOMS2005, Hobart June 2005

[4] Booth, JA & Kautzmann, DE (1984) *Estimating the leakage from HP to IP turbine sections* EPRI Plant Performance Monitoring Conference.

[5] Cotton, KC (1993) *Evaluating and improving steam turbine performance* Cotton Fact Inc ISBN 0 963 995502

[6] Haynes, CJ; Medina, CA; Fitzgerald, MA (1995) *The measurement of HP-IP leakage flow: the largest source of uncertainty in code tests of low pressure turbines* PWR-Vol.28 1995 IEEE-ASME Joint Power Generation Conference

[7] Hogg, S and Stephen D (2003) *Development of Brush Seal Technology for Steam Turbine Retrofit Applications* IJPGC2003- 40103 International Joint Power Generation Conference, June, Atlanta

[8] Hogg, S and Stephen D (2005) *ALSTOM-CPS San Antonio retrofit of JK Spruce Unit 1 HP-IP turbine – an example of an advanced steam turbine upgrade for improved performance by a non-OEM supplier* Proceedings of ASME PWR 2005, April, Chicago PWR2005-50227

[9] Koch, JD; Carriere, DD; Haynes, CJ and Fitzgerald, MA (1988) *Demonstration of Booth's alternative method for measurement of N2 packing leakage* 1988 EPRI Heat Rate Improvement Conference

[10] McCloskey, TH; Dooley, RB; McNaughton, WP (1999) Turbine *Steam Path Damage: Theory and Practice* Vols 1 & 2 EPRI, Palo Alto

[11] Staggers, GA and Priestley, RR (1990) *HP to IP turbine leakage flow measurement: a comparison* ASME/IEEE International Joint Power Generation Conference

STUDY OF REMOTE CONDITION MONITORING AND ASSESSING ON QUAYSIDE-CONTAINER-CRANES

Wang Zhixin [a] Hu Xiong [b] Chen Zhaoneng [a]

[a] *School of Mechanical Engineering, Shanghai JiaoTong University, Shanghai 200030,P.R. China.*

[b] *Department of Mechanical Engineering, Shanghai Maritime University, Shanghai 200135, P.R. China*

Abstract: It is difficult to monitor and assess the health conditions of the Quayside-Container-Cranes (QCCS) on the spot because of QCCs' dispersedly distributions on the docks and bad work environments. In this paper, the techniques of remote monitoring and assessing on QCCs' mechanical conditions are discussed and the corresponding system is developed. The system mainly focuses on monitoring on the stresses of the main metal structures, the vibrations of driving systems and the vibrations of rails and booms. Finally, the system realizes the remote on-line monitoring and remote assessment based on Web.

Key Words: condition monitoring, Quayside Container Crane, remote technique, MATLAB

1 INTRODUCTION

With the rapid development of container transportations, more and more quayside container cranes (QCCs) are put to use. Meanwhile, the QCC's structure gets softer and the vibration and impulsion get more serious because of its bigger size, higher driving speed and heavier hoist loading. So monitoring and assessing their health conditions efficiently play important roles in the plant managements of ports [1].

However, it is difficult to monitor and assess all of QCCs simultaneously, for their disperse distributions and bad work environments. Recently, the technology of remote monitoring makes a remarkable progress with the developments of computer science, communication technique, Internet and WWW. It was firstly implemented in the field of medical treatment [2] and was applied in the engineering field in the 90s of last century [3][4][5]. Now, the remote monitoring and assessing has become a new integrated technology.

The paper does some researches on the techniques of remote online condition monitoring and assessment based on Web, and develops a system that realizes the monitoring and assessment on the stresses of the main structures, the vibrations on the driving systems and the vibrations of the booms and rails of QCCs using the platform of MATLAB.

2 SYSTEM PROJECT

2.1 Monitoring items

Engineering practices show that the multi-parameter analysis makes the condition monitoring more efficiently and assessing more reasonably. What's more, processing data quickly is also very important to remote online monitoring. So it is of great benefit to select the monitoring items that cover more useful information and include fewer parameters.

22 monitoring items are selected and separated into 3 groups after several experiments on QCCs. (the list of items can be seen in Table 1). Group I is defined for monitoring the vibrations of motors and their gearboxes (shown in Fig. 2 and Fig. 3); group II is for monitoring the vibrations of the rails and the nearby parts on booms or beams (shown in Fig.1); group III is for monitoring the stresses of the metal structures (shown in Fig.1).

The mechanic conditions of measuring points on QCCs are denoted as the characteristic values of measurements of sensors. The vibration of the driving system, rail and boom is denoted as vibration severity (mm/s), and the strain of metal structure is denoted as stress (MPa).

2.2 System main tasks

The system realizes two kinds of functions: remote online monitoring and condition assessment based on Web.

For remote monitoring, the system demands that all of QCCs must be monitored simultaneously in the control room that is away from docks. During the monitoring, all characteristic values must be displayed and preserved in the database, and give alarms while abnormal conditions are detected.

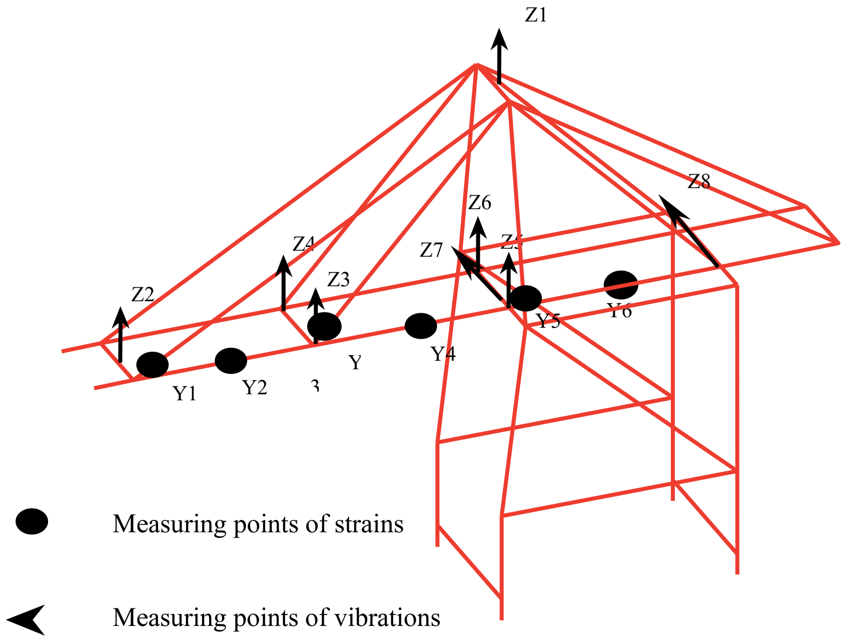


Fig.1 Measuring points on the structure of QCC

Based on monitoring data, the function of assessment is accomplished by using Browser (such as Microsoft IE) through Internet/Intranet. The assessment includes condition distribution statistic, condition forecasting and exception analyzing etc.

The paper constructs the main framework of the system by MATLAB for its powerful abilities of computations. Meanwhile, the paper utilizes the Java, MATLAB Web Server [6] to make up its shortages of network functions. Finally, the remote QCC's condition monitoring and assessment system based on Intranet/Internet are developed.

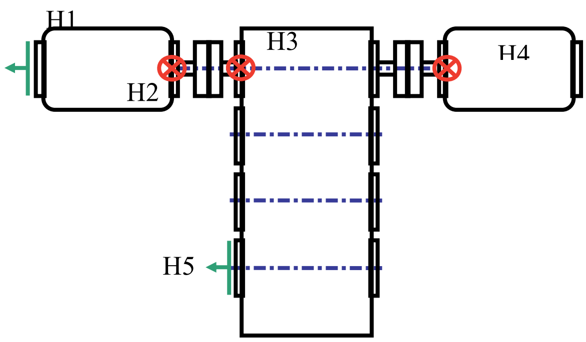


Figure.2. Measuring points on hoisting driving

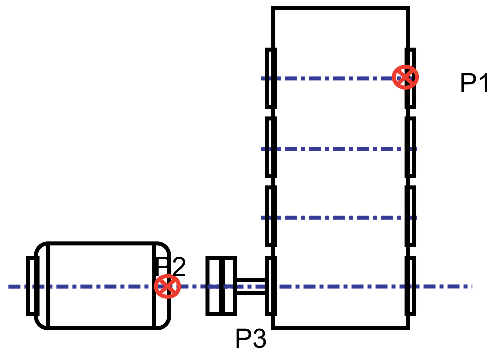


Figure 3. Measuring points on trolley motors and gearbox

3 SYSTEM CONFIGURATION

3.1 System structure

Generally speaking, traditional remote systems adopt Client/Server (C/S) Model. This model exchange Information between clients and servers strictly through messages, so the responses are very quickly, which suits online monitoring very well. In addition, the applications based on this model can accomplish complicated functions, which satisfies the needs of users individually. In the same way, Browser/Server (B/S) Model has been one of the most popular models recently, for its three-layer structure: Client, Web Server and Database Server that divides the functions of application system definitely and independently. The work can be accomplished through browser without installing any client softwares, which avoids many troubles of installation and maintenance.

In this paper, the system integrates advantages of B/S and C/S model. In the remote monitoring subsystem, C/S model meets the need of data transferring and remote monitoring; and B/S model provides a new way for the engineers to assess and predict the conditions of QCCs based on web. Consequently, the framework (shown in Fig 4) are divided into three parts according to the functions and installed positions: RS_R (at the dock); RS_H (in the control room); RS_B (in the office).

Table 1: List of monitoring items

No		Explanations
I	H1	Axial vibration of free end of hoist motor (L)
	H2	Radial vibration of output end of hoist motor
	H3	Radial vibration of high speed axis of hoist gearbox
	H4	Radial vibration of output end of hoist motor (R)
	H5	Axial vibration of low speed axis of hoist gearbox
	P1	Radial vibration of axis of trolley gearbox
	P2	Radial vibration of output end of trolley motor
	P3	Axial vibration of output end of trolley motor
II	Z1	Oscillation in vertical direction of trapeziform girder
	Z2	Oscillation in vertical direction of sinciput boom
	Z3	Vibration of rail and nearby parts on front boom
	Z4	Vibration of rail and nearby parts on middle Boom
	Z5	Vibration of rail and nearby parts at left main joint
	Z6	Vibration of rail and nearby parts at right main joint
	Z7	Vibration of nearby parts at seaside beam
	Z8	Vibration of nearby parts at landside beam
III	Y1	Stress of first pole
	Y2	Stress of front part of the boom
	Y3	Stress of second pole
	Y4	Stress of middle part of the boom
	Y5	Stress of trapeziform girder
	Y6	Stress of middle part of the back boom

3.1.1 Dock:

The RS_R is installed on every QCC. It samples data of the mechanic conditions by means of the sensor equipments installed on every QCC, and translates monitoring data according to the requisitions of RS_H. Since the amount of crude data is usually very large, it is impossible to translate them to RS_H through network in a very short time to realize online monitoring. Only the characteristic values of these data are translated to RS_H, and the crude data are saved in the local real-time database. Simultaneously, RS_S listens connection requests of RS_H, and executes the corresponding actions after receiving requests from the RS_H.

3.1.2 Control Room:

RS_H provides the interface to set the parameters and monitor the health conditions of all QCCs. When engineers chose any QCC to observe the current conditions, the connect request will send to the chosen RS_R to get the real-time condition data. However, the un-chosen QCC are still sampling data synchronously and listening the connect requests from the RS_H.

The monitoring data are saved in the database while monitoring; meanwhile, the crude data at all RS_S are translated to the database periodically. These data are provided to condition assessing, forecasting and other analysis in the future.

The MATLAB Web Server is installed in the RS_H that mainly answers for web server for remote assessment in the office.

3.1.3 Office

Engineers in the office can assess the condition of QCCs by means of web browser (such as Microsoft IE). The RS_B needs none client-software and all works can be accomplished by the server and the results send back to the browser through Internet/Intranet. In addition, The RS_B can be embedded into the plant management of information system, which enhances the automatizations of plant managementTo achieve optimal reproduction quality, ensure that the contrast of text lettering is uniform, sharp, and dark over the whole page and throughout the paper. If corrections are made to the printout, run-off completely new replacement pages. The contrast on these pages should be consistent with the rest of the manuscript as should text dimensions and font sizes.

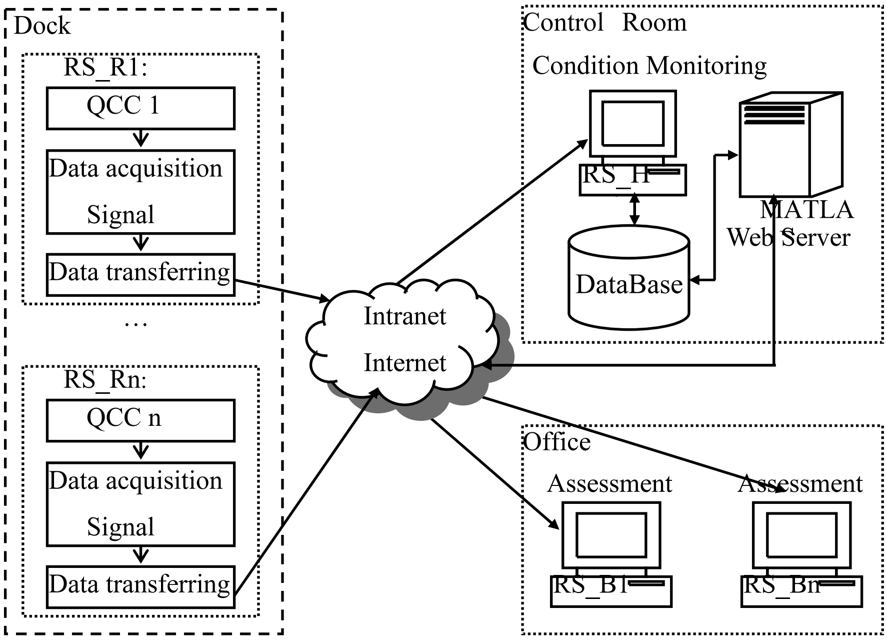


Figure 4. Structure of the system

3.2 Remote Data Transferring

The primary problem in the remote system is data translating. It needs translating the condition data between RS_R and RS_H timely while real monitoring and send parameters to RS_R precisely. Unfortunately, MATLAB lacks for the network functions and cannot support Sockets. Nevertheless, Java provides an integrated set of class libraries for network that very suit the development of remote monitoring systems in distributed measurements [8]. Users can design their special class libraries based on them, which realizes the Java's characteristics expediently. These special libraries install to MATLAB to help MATLAB to support network functions. Subsequently, two kinds of libraries for Server and Client are designed individually according to the special acquirements:

1) Class libraries in RS_R listen the requests of RS_H and translate data between each other (The detailed flow is shown in figure 5). Class HostSer is developed based on the class ServerSocket. When a new object of HostSer is created, it begins to listen possible requests from RS_H at the special port and then offers services such as sending files; accepting files; deleting files while connecting. After the connection is closed, the RS_R recurs to listens status.

2) Class libraries in RS_H are divided into five according to their functions (listed in the figure 5) that answer for accepting data from RS_R, sending data to RS_R, obtaining file lists of the appointed directory, deleting the appointed files and closing the communication individually. The five libraries are developed by the Socket class. All programs that want to communicate with the each other by means of Sockets must create objects of the Socket class.

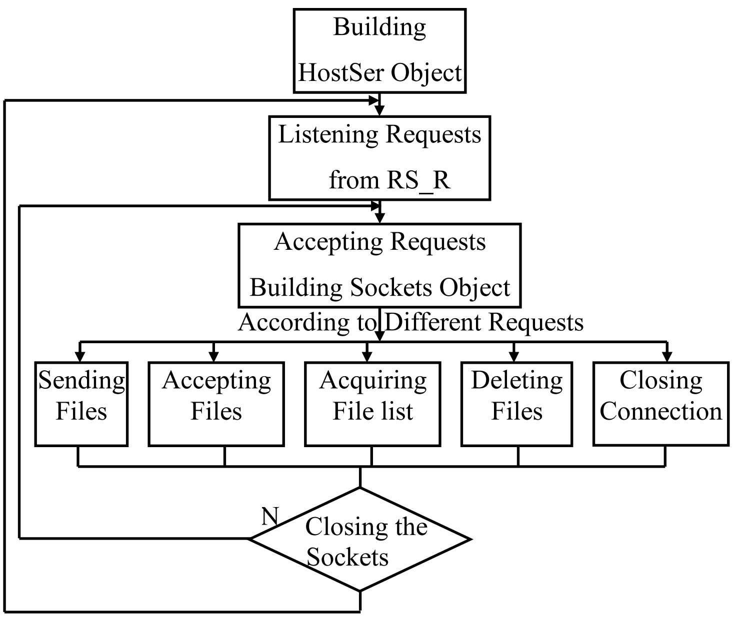


Figure 5. Flow of Library of HostSer

Further more, the network blocking must not disturb the course of MATLAB in gear. Meantime, there may be more than one RS_H that ask for connections to RS_R. Therefore, programs need the functions of multi-client connections. Based on the characteristics of Java, the system realizes the function of multi-thread that explained in figure 6 in detail

3.3 MATLAB Web Server [7]

The Web application of MATLAB is composed by two parts: one is the MATLAB Web Server (matlabserver.exe) which manages the communication between the Web application and MATLAB; the other is the Web server proxy which uses the Common Gateway Interface (CGI) to extract data from HTML documents and transfer it to matlabserver, then send back the results in manner of HTML (The basic theory shown in Fig. 7).

In this system, the MATLAB Web Server and the Web server daemon (httpd) are installed in the control room together with the database. When Users assess the conditions in the office, they send requirements to server through browser (Microsoft IE); and the server answers for looking for data in the databases, analyzing, and returning the results back to the clients through WebPages

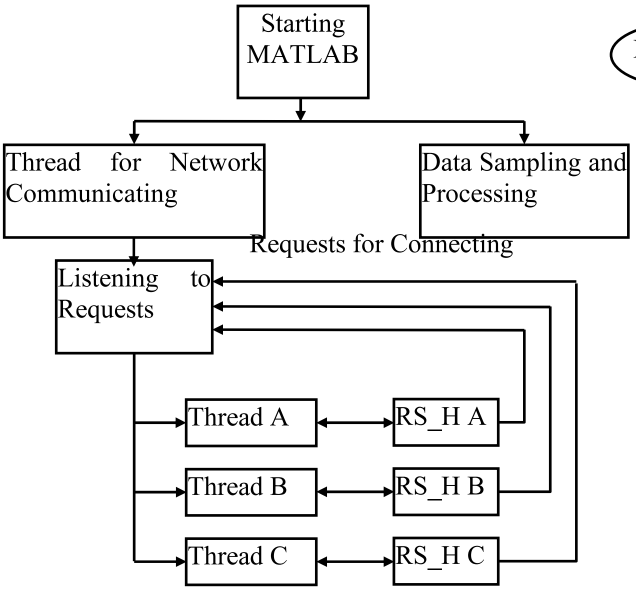


Figure.6 Principal of Multi-thread at RS_R

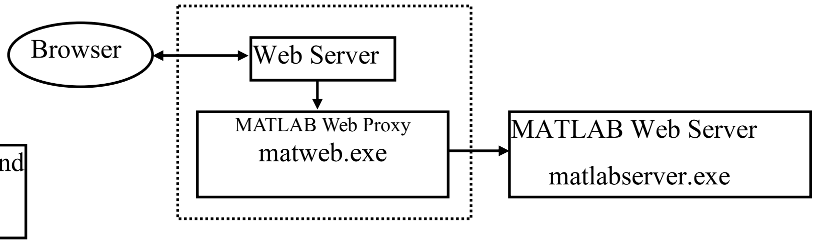


Figure 7. Basic theory of MATLAB Web Server

4 ERFORMANCES AND CONCLUSIONS

Figure 8 shows the interface of RS_H where users set parameters and monitor the conditions of QCCs. Users choice the number of QCCs and the group of monitoring items at the top of the window. Users can set different levels above which the crude data will be preserved at RS_R. These data can be used to do further researches.

When the system operates, users get the real-time data at intervals of about 10 seconds. The data of the chosen group are shown in the form of chart that moves with time, and the digitals of all items are displayed at the bottom at the current time, and the alarm lines of different levels displays in different colors.

Function of "Condition Assessment" emphasizes the condition distributions in a special period of time. The conditions are divided into 5 levels (A, B, C, D and E). Because the work environments are very complex (for instance, braking suddenly may cause violent impulsive vibrations), a few bad conditions caught cannot say there is malfunction on QCCs. Statistical analysis can makes the result more reliably and believably. Once the distributions of these conditions change, the conditions of the QCC should be analyzed in detail. Figure 9 shows the distributions of conditions of 22 items from 2005-03-07-10'55" to 2005-03-07-12'41" that covers 839 pieces of data. The green line indicates the distribution of real data while the red indicates the normal distribution whose variance and mean are equal to those of the real data, and the condition level distributions are listed in the page.

This thesis does some researches in remote online condition monitoring and assessment based on web. Some class libraries are built which enriches the network functions of MATLAB based on Sockets; and the technique of MATLAB Web Server helps to realize remote assessment based on Web.The monitoring and Assessment System is then developed to online monitor and assess to the mechanical conditions of the QCCs' main parts. Many experiments have been made to test its reliability and veracity. Now, it has been utilized in the practical operations, which enhances the efficiencies of plant management of QCCs.

5 REFERENCES

[1] Hu Xiong, Xu Dan, Li Yong (2002). System Design for Mechanical Condition Monitoring and Safety Assessment Of Quayside Container Crane. Proceedings of the 3rd Asia Pacific Conference on Systems Integrity and Maintenance.

Figure 8 Interface of RS_H

Figure 9 Web page of Condition Assessment

[2] Tsuij Sadatoshi, Akamatsu NaokiMurai, Yoshiyuki Tobimatsu, Shozo Kato Matohito (1997). Remote Diagnosis of intractable epilepsy by bi-directional tele-medicine system between USA and Japan. Electroencephalography and clinical Neurophysiology. Vol.133 Issue: 1.

[3] Jim M N, Tsang P H H, Tsang C H K (1994). Network Requirements for Remote Diagnosis Consultation [J], Industrial Technology.

[4] Boning D, Chen K N (1997). Next Generation Remote Monitoring and Diagnosis [C], LFM Symposium.

[5] U.Kunze (2003), Condition telemonitoring and diagnosis of power plants using web technology. Progress in nuclear energy. Vol.43, No.1-4, pp.129-136.

[6] Zhao Aiping (2001). Matlab Web Server and its application in remote collaborative design of magnetic bearing systems [J]. Chinese Journal of Mechanical Engineering, 14 (2): 179 - 183.

[7] MathWorks Inc (1999). Matlab Web Server, The Language of Technical Computing. http://www. mathworks. Com..

[8] L. Reveilleau, E. Becquet (2001), Towards a RT_Java Based Embeded Remote Monitoring Tool for small and Medium Power Plant Units. IEEE.

6 ACKNOWLEDGMENTS

The authors would like to thank the Shanghai's Education Commission for financially supporting this research under Contract No. 2004095.

DETERMINATION OF THE RISK PROFILE OF A HYDROGEN CONSUMING FACILITY USING PUBLISHED FAILURE RATE DATA

Thomas I F

Principal, I F Thomas & Associates Pty Ltd, Consulting Risk Engineers

Abstract: This paper assesses the potential consequences and associated risk levels for a major explosion following catastrophic failure of an hypothetical hydrogen-containing plant operating at elevated temperature and pressure. The plant is mooted as being located in an industrial area but with residential areas nearby. It considers the importance of fully open-air construction of such plants, of sufficient separation from surrounding communities and of free vertical discharge from pressure relief devices. It poses the question that the authorities in Australia may have got it wrong in not designating large hydrogen gas users as major hazard facilities with all that this entails. With the aid of a series of cartoons the paper brings us back from predictive mathematics to reality by indicating that common sense can be better at preventing accidents than even the best quantified risk assessment - do the risk assessment but do not rely upon it blindly. The paper concludes that an explosion of sufficient size to seriously embarrass or even close-down a major hydrogen gas using facility is possible. Likelihood of this is capable of being reduced to a negligible level by good design and by conducting a detailed risk assessment ahead of acceptance of the design. The magnitude of an explosion should such a large leak occur would also be reduced by the buoyancy effect of hydrogen, provided that the plant is constructed in the open air.

The base level of individual risk associated with catastrophic failure of the hydrogen-containing facility is established as 189.2 x 10^{-6}/person/year. This level of risk would be unacceptable if the 50% mortality distance extended into the community surrounding such a facility. The true level of individual risk associated with the same catastrophic failure but considering all of the mitigation measures inherent in the facility, is 1.78 x 10^{-84}/person/year. The latter 'risk' level is purely theoretical and shows that the failure modelled is essentially impossible. It measures the event frequency per unit of time akin to the Googolplex *(10^{100})* and to the purported number of particles in the universe *(10^{80})*. Risk levels have numerical meaning only if they are measured in units of powers of ten ranging from 0 to minus 10 or so and are effectively irrelevant once shown to be less than the prevailing criteria.

Key Words: Risk, Hydrogen, Explosion, Vapour cloud explosion, QRA, Buoyancy, Failure rate Format

1 THE HYPOTHETICAL NEIGHBOURHOOD

The property is assumed to be located wholly within an Industrial 1 Zone with most immediate neighbouring properties similarly zoned. Exceptions are public parks and nearby municipal premises. Within 500m of the centre of the site are more sensitive uses including small businesses, a busy freeway and residences. A railway line is within 500m of the centre of the premises. A dense residential zoning including some tall buildings is located beyond 500m. The latter are technically of the same degree of sensitivity as single-storey residences but contain more residents per unit of land area and large expanses of glazing. Hence the premises can be described as being ideally located in terms of immediate neighbours, less ideally with respect to impacts which may extend to 500m and even less ideally beyond 500m.

Key to whether the location may lead to adverse impacts upon its neighbours, is the extent to which these impacts may extend out into the community beyond site boundaries in particular, as far as residences. The reason for singling out residences over and above industry, businesses, road, transport and public park zones is because residences are deemed to be occupied continuously, 24 hours per day, 365 days per year. Workplaces in the form of industry and local businesses are deemed less sensitive in that they are occupied for periods per day less than this, typically from 8 to 12 hours per day. Sensitive areas such as sports grounds and open parks may be densely occupied but again, for considerably less than 24 hours per day. Classically therefore, risks of exposure to harm are orders of magnitude less for less frequented zonings. Occupants at the subject premises are considered as a special case in risk terms, because they are either in control of potentially adverse events or in immediate communication with those who are.

2 RISK EVALUATION PROCESSES AVAILABLE

2.1 Think tanks, fault tree analysis and HAZOP Studies

The facility is assumed to employ full time engineers and a safety officer and therefore to competently address both day-to-day risks and less likely, higher consequence risks. The company's engineering system is assumed to use to great effect Hazard and Operability studies *(HAZOP),* Job Safety Analysis *(JSA),* Safe Work Practices *(SWP),* hot and cold work permits and confined space entry permits. All works and pertinent documentation are assumed to be rigourously filed.

2.2 Notional (qualitative) risk assessment

The authorities almost exclusively require occupiers of potentially dangerous premises to conduct a documented notional risk assessment. This is generally conducted in worksheet or spreadsheet format, and considers adverse events called hazards, the consequence and likelihood of these hazards and thereby, the level of risk associated with them. The same or a separate worksheet then considers available risk mitigation measures and develops a program of works for addressing risks, concentrating on those hazards which present the highest risk. The use of worksheet style is beneficial in permitting presence of rows addressing each hazard and columns stating for example, the item, hazard, consequence, severity of consequence, likelihood of occurrence, risk level, mitigation and whether new action is warranted. An example of such a 'generic' risk assessment is appended.

Risk in these assessments is best provided with a number so that relative risk can be established and priorities set for attending to them. AS/NZS 4360:1999, Risk Management, addresses this. Notional risk assessment which arrives at a series of risk numbers is sometimes referred to as semi-quantitative or semi-quantified risk assessment. The example of a risk matrix shown in the standard, uses a scale of 1-5 for severity of consequences, A-E for likelihood and consequent risk levels described as E *(extreme),* H *(high),* M *(moderate)* and L *(low).* A technique considered to be more useful assigns levels on a 1-5 scale to both severity of consequence *(S)* and likelihood of occurrence *(L),* thereby permitting Risk *(= S x L)* to be established as one of 14 numbers in the range 1-25. Priority risk levels of eg 10 and above are then easily recognised.

2.3 Quantified (or Quantitative) Risk Assessment (QRA)

This is a sophisticated technique involving relatively complex mathematics to provide absolute risk levels for a defined event or series of defined events. It uses inputs such as actual frequencies of occurrence and failure rates together with analysis of probabilities. Numerous computerised models exist and are offered by various companies on a commercial basis. Models exist for overall QRA of a premises and for specific events such as an escape of flammable gas or vapour. Computer models are complex and usually proprietary. Even licenced use is restricted and often may not permit the user to know the mathematics which are employed. Outputs are contours drawn in plan around a premises, showing the **individual risk of death** per person per year. A single event in a non-complex plant on flat terrain without obstacles will yield a series of circles each representing a different risk number. They are known as contours of individual risk. Risks other than death can also be assessed but it is the risk of death which dominates these studies. A common acceptable individual risk of death contour is 1×10^{-7} *per person per year*. Levels less than this are generally considered acceptable. It is evident that the risk level is defined both by the event and by the density and degree of presence of the surrounding population at distances and locations specified.

A second and equally important output from QRA is known as the group or societal risk. This considers catastrophic events capable of causing multiple fatalities. This is not shown as contours but as a graph showing the number of deaths (N) resulting from an event which occurs at a frequency (F). This is known as an F/N plot and again, a range of acceptable levels of societal risk is published by a number of authorities. An **acceptable risk range** published by one of the authorities shows *negligible risk below a line bounded by* $N=1;\ F=10^{-4}$ *and* $N=30;\ F=10^{-7}$. **Unacceptable risk** is defined as being *above a line bounded by* $N=1;\ F=10^{-2}$ *and* $N=300;\ F=10^{-7}$. The **region between these two lines** is deemed an area of **acceptable risk but where remedial measures are desirable**. This F/N plot is appended hereto.

Quantified Risk Assessment in the 1980's and 1990's was considered by the authorities as the ultimate way of assessing risk and use of QRA was vigourously supported. Authorities even purchased their own models to independently consider the risks of particular premises. Authorities have defined and published acceptable risk levels with which the outputs of QRA were to be compared and actions developed there-following. However, examples exist where authorities have accepted the risk levels calculated for a premises even though they were outside their own criteria. QRA also gradually acquired the reputation of being 'malleable' to fit with a desired outcome. Even current, highly sophisticated models are best only used for comparison of one facility with another or of the same facility at different times where the same model was used in each case. In one instance, a government approved a new development proposal which was shown to cause individual risk levels above the criteria in nearby residential areas. In another, a QRA showed a new plant proposal involving installation of bulk monomer storage tanks close to residential neighbours. Objections led to a second QRA being performed which showed that individual risk levels increased if the proposed tanks were not installed.

2.4 The method used herein

A somewhat conservative assessment is conducted herein, involving assessing the maximum possible *(albeit unlikely)* hydrogen gas escape, converting this to an equivalent mass of the conventional explosive trinitrotoluene *(TNT)* and then establishing effects with distance of a TNT explosion of that size consequent upon explosion *(or blast)* overpressure. The technique does not take account of terrain, obstacles, buoyancy, wind, atmospheric dispersion conditions or the effects of falling structures and flying debris. Rather, it assumes flat terrain, neutral buoyancy, no wind, absence of obstacles and that there are no injuries, deaths or property damage resulting from falling structures and flying objects and does not consider the impact of atmospheric dispersion conditions. Probabilities are assigned in the *Likelihood* section hereof. The technique also assumes that an escape of hydrogen gas is similar to the escape of a heavier-than-air gas or vapour like ethylene, vinyl chloride or cyclohexane.

The method herein relies upon published works by Gugan, Bond, Brasie & Simpson, Lawrence & Johnson, Strehlow & Baker, Clancy, Marshall and Sutton. Actual references are listed in the appendix hereto. The effects of explosions with distance are defined by overpressure present at that distance. Herein, the method of Lawrence and Johnson *(1974)* is used to predict human injury effects. These are described below together with the associated overpressure :-

50% mortality *20psi =138kPa*
Lung damage *5psi = 34kPa*
Ear damage *2.5psi = 17kPa*
Penetration by 10g missiles *0.4psi = 3kPa*
Cuts by flying glass *0.2psi = 1.4kPa*

The method of Strehlow and Baker (1976) is used to predict property damage for various levels of overpressure as follows :-

Total destruction *10psi = 69kPa*
Serious damage *3psi = 21kPa*
Minor structural damage *0.4psi = 3kPa*
Glass breakage *0.03psi = 0.2kPa*

Both groups of workers agree that glass typically fails at 0.1-0.2psi *(0.7-1.3kPa)* but as a conservative measure, the *limit* of glass breakage as opposed to the typical overpressure to cause it is used namely 0.03psi *(0.2kPa).* It should be noted that glass breakage caused by outward showering of broken glass from nearer to the explosion, extends out considerably further than the pressure wave which is capable of breaking glass at a distance.

3 CALCULATIONS

3.1 Establishment of worst-case hydrogen release

This assumes that the vessel containing hydrogen is filled with 20,000L of hydrogen at a pressure of 2200kPa and at a temperature of 10ºC. Whilst essentially impossible as this would require major failings in safeguards, it is then assumed that the reactor ruptures and releases all of this hydrogen essentially instantaneously.

Using the ideal gas law, $PV = nRT$ we have
$P1$ = 2200kPa
$V1$ = 20,000L
$T1$ = 10ºC = 283ºK

Converting to Standard Temperature and Pressure *(STP)* we have
$P2$ = 101.3kPa
$V2$ = V2
$T2$ = 0ºC

$P1V1/T1 = P2V2/T2$

Therefore $V2 = 2200 \times 20{,}000 \times 273/283 \times 101.3 = 419{,}005L$ of hydrogen

Using Avogadro's Hypothesis, *1 gram mole of an ideal gas at STP occupies 22.4L*
Therefore weight of hydrogen = *419,005 x 2/22.4 = 37,411g = 37.4kg*

As this study is intentionally extremely conservative, we will proceed using 40kg of hydrogen.

Heat of combustion of hydrogen = *33,887.6kcal/kg*

Heat of combustion of TNT = *1100kcal/kg*

TNT equivalent of 40kg of hydrogen = *40 x 33887.6/1100* = *1232.3kg*

Hydrogen is considerably more prone to explosion *(detonation)* than rapid burning *(deflagration)* when compared with heavier-than-air gases and vapours. The range of *blast wave efficiencies* used by various workers is from 2% to 30%. Again because this appraisal is conservative and because hydrogen is more likely to detonate than to deflagrate *(detonation range in air = 18.3-59%; flammability range in air = 4-75%; ignition energy = 0.01mJ),* a blast wave efficiency of 30% is assumed, yielding

TNT equivalent explosive power = *1232.3 x 30/100*

= *369.7kg*
= 0.3697t of TNT (W)

The pressure wave from a TNT explosion of this magnitude will cause human injury and property damage, the effects decreasing with distance. These are expressed as distances *(d)* up to which certain effects are expected.

3.2 Human injury distances

50% mortality d $= 17.8W \times 10^{0.33}$ = **12.8m**

Lung damage d $= 35.8W \times 10^{0.32}$ = **26.0m**

Ear damage d $= 53.8W \times 10^{0.32}$ = **39.1m**

Penetration by 10g missiles d $= 316.2W \times 10^{0.33}$ = **227.7m**

Cuts by flying glass d $= 640.4W \times 10^{0.32}$ = **465.6m**

3.3 Property damage distances

Total destruction	d	= 4W x 100.33	= 2.9m
Serious damage	d	= 10W x 100.33	= 7.2m
Minor structural damage	d	= 30W x 100.33	= 21.6m
Glass breakage	d	= 430W x 100.33	= 309.6m

3.4 Total deaths

Total deaths can be estimated using criteria developed by Marshall and by Gugan as follows :-

Total deaths	= 0.05W x $10^{0.33}$	= 0.04 (considered as 0)
Total deaths	= 0.3W x $10^{0.33}$	= 0.2 (considered as 1)
Total deaths	= 4W x $10^{0.5}$	= 2.4 (considered as 3)
Total deaths	= 3.3W x $10^{0.33}$	= 2.4 (considered as 3)
Total deaths	= 55W x $10^{0.5}$	= 33.4 (considered as 34)

Predictions of **total deaths** for an explosion involving **0.3697t of TNT** therefore range from **0 to 34**.

3.5 Other scenarios considered

- A 20,000L charge of hydrogen at 1700kPa and 250ºC ***(15.6kg of hydrogen)***
- Vessel filled with 20,000L of hydrogen at 500kPa ***(8.8kg of hydrogen)***
- Vessel filled with other materials and containing hydrogen at 2200kPa ***(3.9kg of hydrogen)***

Each of these are lower weights of released hydrogen than the case pursued above. Results for each follow :-

Hydrogen released (kg)	*3.9*	*8.8*	*15.6*	***40.0***
Human injury distances (m)				
50% mortality	*5.9*	*7.8*	*9.4*	***12.8***
Lung damage	*12.4*	*16.0*	*19.3*	***26.0***
Ear damage	*18.6*	*24.1*	*28.9*	***39.1***
Penetration by 10g missiles	*105.6*	*138.1*	*166.9*	***227.7***
Cuts by flying glass	*221.0*	*286.8*	*344.6*	***465.6***
Total deaths range	*0 to 10*	*0 to 16*	*0 to 21*	***0 to 34***

Property damage distances (m)

Total destruction	1.3	1.7	2.1	**2.9**	
Serious damage	3.3	4.4	5.3	**7.2**	
Minor structural damage	10.0	13.1	15.8	**21.6**	
Glass breakage		143.6	187.8	227.0	**309.6**

The easiest way to consider the significance of distances established in the foregoing is by use of maps and plans where the effect distance 'd' is drawn as a radius centred on the location of the facility. In a purely analytical way, it is fair to suggest that effects which are confined within the facility are less devastating than off-site effects. Serious effects like death, lung damage, ear damage, total destruction, serious damage and minor structural damage in the worst case considered here, are all confined within the site. It is only lesser effects such as penetration of the body by 10g missiles, cuts from flying glass and glass breakage which extend off site. Whilst causing injury of any kind off site would lead to actions potentially resulting in plant closure, this work is purposely an exaggeration of what might actually happen.

4 THE EFFECTS OF VARIABLES NOT CONSIDERED

4.1 Wind

Wind has a more powerful effect in mitigating the extent of the consequences of an accidental release than is generally realised. The writer has developed equations indicating this for heavier-than-air releases, viz :-

$Wvc = 794.3\ u^{-0.87}$
$Wtnt = 1413\ u^{-0.85}$ where

Wvc is the weight of the leaked substance *(tonnes)*
Wtnt is the equivalent weight of trinitrotoluene &
u is the windspeed *(m/hr)*

Even a small breeze can disperse an otherwise flammable cloud to below flammable limits either quickly in the event of a single release, or at relatively close distances to the source in the event of a continuous release. An example of the latter is the B F Goodrich Chemical Ltd, Altona release of 676t of vinyl chloride over a period of 5.25hrs in 1982. Average emission rate was 2.15t/minute and windspeed was a very moderate 8-14 knots. It was calculated that at no time was more than 150kg of vinyl chloride within the flammable envelope. With complete absence of wind it was shown that more than 100t would have been within the flammable vapour cloud during the steady-state emission phase.

4.2 Emission rate

$Wvc = 0.6 \times 10^{-6} \times Qm$

$Wtnt = 1.1 \times 10^{-6} \times Qm$ where

Qm is the mass emission rate of the release (kg/hr)

4.3 Atmospheric dispersion conditions

The condition of the atmosphere ranges from highly dispersive in an upward direction to providing a 'lid' and preventing upward dispersion, depending upon atmospheric conditions and time of day. This is more important for heavier-than-air emissions than for a substance like hydrogen which is buoyant regardless of weather conditions. Atmospheric dispersion condition nevertheless is capable of slowing-down or speeding-up the dilution of a cloud of hydrogen to below flammable limits.

$Wvc = 7.41D^{-0.51}$

$Wtnt = 12.58D^{-0.5}$ where

D is the dispersion parameter = $m \times n^2/4$, where
$D = 100Cy.Cz/n$ and where *Cy* and *Cz* are vertical diffusion parameters and *n* is a dimensionless turbulence parameter.

4.4 Terrain

The relatively flat terrain of the hypothetical site and its surroundings means that this effect need not be considered. Clearly, flat terrain will lead to greater dispersion than an instance of an emission in a valley and vice versa. Most computer models also incorporate a surface roughness factor.

4.5 Obstacles

Presence of obstacles such as walls, roofs, plant, buildings and houses will influence dispersion of a leak and will also more importantly, increase the likelihood of a cloud catching fire and of detonating. Open air emissions until the 1970's were generally regarded as harmless deflagrations. It is now understood that presence of obstacles can convert a deflagrating cloud into an explosion. It is noteworthy that a vapour cloud explosion generally causes overpressure for longer than conventional explosives such as TNT. Structures designed to withstand TNT may therefore fail in a vapour cloud explosion. Increasingly, vapour cloud explosions are being implicated in major events such as Flixborough UK *(1974)* and more recently Buncefied UK *(December 2005).*

5 HYDROGEN AS OPPOSED TO HEAVIER-THAN-AIR EMISSIONS

Hydrogen gas has a very high flammability range *(4-75% in air)* and a high detonation range *(18.3-59% in air).* Together with its extremely low ignition energy *(0.011mJ @ 25ºC; 0.0051mJ @ 150ºC),* almost any release will ignite and most will detonate. Diffusion of hydrogen leaking from a pressure source will of itself, cause ignition if the pressure source is at 7900kPa or greater. Hydrogen is typically stored at 18,000kPa. Reactor pressure is in the range 400-2200kPa. Most emissions of hydrogen with which the author is familiar have resulted in detonation albeit that these have all been small for example, from G-sized gas cylinders and from continuously operating electroplating baths. Shorting of vehicle battery terminals also often results in a minor explosion. One instance is known where the battery exploded sending components and acid several metres.

Information to the contrary has been obtained pertaining to use of hydrogen balloons to float cameras inside caves. Here, the process of filling the collapsed balloon bags was described as being plagued by sparks which did not cause ignition. Disposal of such a balloon over fire has also been described as a relatively harmless, non-exciting event. However, accounts like the latter need to be tempered with the reality as first discovered by Pilatre de Rozier in the 1700s when he repeatedly demonstrated the safety of hydrogen by inhaling it and setting fire to his breath via a glass tube. On one occasion an explosion resulted from which he survived. In 1784, de Rozier went further by developing a lifting device combining a hydrogen balloon with a hot-air balloon. This time he was killed.

It has been a common misconception over a long period of time, that the infamous crashes of airships such as the Hindenburg and R101 were caused by hydrogen catching fire. This is not the case and resulted in the utilisation of hydrogen and of airships being set-back for many decades. Confined hydrogen can cause serious pressure escalations. Unconfined, a simple combustion or deflagration will usually occur resulting in overpressure up to 7kPa. A confined deflagration can raise this pressure 8-fold whilst a detonation can raise the pressure 15-fold. A confined combination of deflagration and detonation however, can lead to pressures rising 120-fold.

Provided the release is unconfined or relatively unconfined, its extreme buoyancy *(it is the lightest element known and is 14.4 times lighter than air)* will cause an emission to move away upwards from source, quickly. Despite the high propensity to ignite and then detonate, this tendency to move away is considered important in any consideration of what might happen in the event of a major release. As is common in pressurised storage, it is therefore desirable for hydrogen processes to be in open-air or at least in buildings with partial walls and partial roofs.

Buoyancy relationships show that hydrogen will accelerate upwards once released at 131m/s/s until reaching its terminal velocity of 17m/s. Hydrogen is the only element capable of reaching escape velocity *(11.2km/s)* but it is understood that this much higher velocity is only achieved in the upper atmosphere. Therefore after only a few seconds following a release, in such an unconfined situation, the gas will be tens of metres into the air and will have already dispersed considerably in so doing. The occurrence of an elevated explosion has the disadvantage of being capable of sending blast waves outwards in the way indicated herein as if no buildings or obstacles were present. Elevated residences would continue to be vulnerable if the explosive power was great enough. On the other hand, an elevated explosion also has the advantages of being more distant from the source when it ignites and being more dispersed and therefore less likely to detonate. At best, dispersion associated with the rapid upward movement will prevent ignition. The extent to which the buoyancy of hydrogen mitigates against major disaster should be capable of being established by application of one of the computer models available and also by reference to past accidents. By using windspeed relationships referred to above, it is clear that a vertical 'windspeed' of 17m/s would greatly reduce the impact of such a release. Assuming that the proposed plant is in the open air, it is reasonable to regard the hydrogen upflow as a vertical wind. This then permits application of the wind effect equation, namely :-

$$Wvc = 794.3\ u \times 10^{-0.87}$$
$$Wtnt = 1413\ u \times 10^{-0.85}$$

Where *Wvc* is the weight of leaked substance *(tonnes), Wtnt* is the equivalent weight of TNT and *u* is the windspeed in m/hr. These relationships were established for an emission rate of 33.3kg/s over a period of some 5 hours *(total release = 600 tonnes).* Taking 'windspeeds' of 17m/s, 1m/s and 1m/hr we find

$$Wvc = 54.4kg\ @\ u = 17m/s$$

$$Wvc = 639.7kg\ @\ u = 1m/s$$

$$Wvc = 794{,}300kg\ @\ u = 1m/hr$$

Windspeeds of 1m/hr or similar are unrealistic as air movement is almost always greater than this and as the weight of vapour in the explosive state exceeds the total emission in the example from which the equations were derived. It is also unrealistic to observe that more hydrogen than that present, is emitted in the 17m/s and even in the 1m/s instance. However, if it is assumed that the ratio of wind-dilution effects is the same for the modelled case and the case which provided the equations, we have

Ratio of windspeed impacts from *17m/s to 1 m/s = 11.76*

Therefore we might assume that an actual *(instantaneous)* release of a total of *40kg* of hydrogen, would reduce to

40/11.76 = 3.4kg consequent upon the presence of a *17m/s* updraft.

This derivation is not claimed to be accurate but directionally, it shows that the scale of an explosion would be reduced significantly by the presence of such an updraft. Hydrogen reaches its terminal velocity in the following distance

$v^2 = u^2 + 2as$, where

v = 17m/s
u = 0m/s
a = 131m/s/s
s = s metres therefore

s = 1.103m

and in the following time

$s = ut + 0.5at^2$ where t = time (seconds)

$1.103 = 0.5 \times 131 \times t^2$ therefore

t = 0.017 seconds

The upward acceleration of hydrogen is thus so large that within a fraction of a second and just over a metre above the source, the terminal velocity is reached. Therefore the dispersive impact of the upflow velocity as herein surmised, is pertinent. If there is no ignition source at point of emission, it is possible that the rising hydrogen will reach terminal velocity without finding an ignition source and that then, the speed of rising will disperse the emission rapidly to below the lower flammable limit *(lower explosive limit, LEL)* concentration. This further indicates the importance of the upflow effect. Actual experience however is that large releases of hydrogen involving significant dropping in pressure, almost always ignite.

6 LIKELIHOOD

This may be established by reference to industry 'failure rate data'. Published pressure vessel failure rates are for example :-

Lees, F P, 3rd Ed, Vol 3, A14/7 (general; UKAEA) :-	Failures/10^6 hrs of operation
Pressure vessels (general)	3.0
Pressure vessels (high standard)	0.3

Lees, F P, 3rd Ed, Vol 3, A7/27 (1st Canvey) :-	Frequency of initiating event x 10^{-6}/yr
LPG explosion due to spontaneous failure of storage sphere (Occidental, UR)	150
Spontaneous failure of LPG vessel (Shell)	140
Spontaneous failure of vessel (Mobil)	230
Spontaneous failure of vessel (Occidental)	90
Spontaneous failure of vessel (UR)	70
Spontaneous failure of vessel (Fisons)	100

Lees, F P, 3rd Ed, Vol 3, A7/31 (2nd Canvey) :-	Frequency of event x 10^{-6}/yr
Spontaneous failure of pressure vessels	10

Lees, F P, 3rd Ed, Vol 3, (Rijnmond) :-	Frequency of event x 10^{-6}/yr
Pressure vessel - serious leakage (A8/6)	10
Pressure vessel - catastrophic failure (A8/6)	1 (range 0.63 to 46)
P1 Major rupture of tank (A8/9)	43
U0 Catastrophic failure of full NH_3 sphere (A8/13)	0.23
U1 Catastrophic failure of part-full NH_3 sphere (A8/13)	1.8
A1.1 Catastrophic burst, chlorine tank, full (A8/14)	0.093
A1.1 Catastrophic burst, chlorine tank, half full (A8/14)	0.74
G1 Catastrophic failure of LNG tank (A8/15)	0.8
0.01 Failure of full propylene sphere (A8/16)	0.028
0.02 Failure of propylene sphere c/w 300t (A8/16)	0.233

AIChE, CCPS Data Tables :-	Failures/10^6 hrs of operation
Vessels, pressurised, metallic (Catastrophic failure, leakage < ¼")	0.0424

Thomas, I F, Frequency of event x 10^{-6}/yr :- (Chemical industry data; unpublished)	
Catastrophic failure of 12t polymeriser	196

The above data vary widely. Results are first converted to events per year assuming 8000hrs/yr of operation. Experience and confidence in high-standard nuclear pressure vessels (as opposed to low standard) leads reasonably to elimination of the 24,000 x 10^{-6}/yr (3.0/10^6 hrs) result. The average of all remaining 20 figures leads to the following

Highest 2400 x 10^{-6}/yr
Lowest 0.028 x 10^{-6}/yr
Average189.2 x 10^{-6}/yr

If this data alone represented the situation at the subject premises, the likelihood of catastrophic failure of the hydrogenation reactor on average, is 189.2 x 10^{-6} per year. This would mean that the individual risk to persons located in the community within the effect radii shown, would be

***189.2 x 10^{-6} /person/year**.*

Acceptance criteria in various countries including Australia are cited earlier herein. A community *individual risk of death* criterion informally accepted in Victoria and formally accepted in many other jurisdictions is

1 x 10^{-7}/person/year

It is shown above that risks of death and of serious injury are confined to the site for both a 40kg explosion and the second modelled, 3.9kg explosion. Therefore strictly speaking, the effects which these events project into the surrounding community would not be sufficient to cause the authorities to refuse an application to build such a plant. However, if such an hypothetical plant was to be built, and if it was capable of causing glass breakage, cuts from flying glass and penetration by 10g missiles in the community as indicated, it is suggested that this would not be an acceptable situation.

Therefore if the likelihood of a catastrophic failure of the hydrogen-using facility was actually as above, this would not be acceptable. However, the foregoing data include many incidents of such failure where a small hole and a harmless jet discharge ensues as well as actual, major disasters. Only two records of hydrogen explosions of any scale have been found in the literature *(Lees F P, Incident B33, Appendix 1/73; CRC, Hydrogen: Its Technology and Implications, 1977)*. In the first, hydrogen was intentionally vented when there was no wind and an explosion followed. In the second, the vessel burst under pressure after 12 years of operation.

7 THE ACTUAL LEVEL OF RISK OF EXPLOSION

To establish the actual failure rate of such an hypothetical facility, the most likely cause of catastrophic plant failure must be considered. This is over-pressurisation from hydrogen storage which is typically at *165bar (16,500kPa).* There are numerous safeguards to prevent this and any other adverse event in the plant and these will be discussed here-following. A simplified Process Flow Diagram *(PFD)* is appended hereto to assist an understanding of these safeguards.

In the single hydrogen supply line from storage to process there are mooted to be some 30 devices comprising manual valves, pneumatic and solenoid control valves, relief devices, pressure transmitters, flow transmitters, alarms, indicators and in addition, a sophisticated computer control system.

All components have levels of reliability enhanced by routine inspections, by preventive maintenance, by good operation, and by observation & diligence of engineering personnel. However for the purpose of this study, each component also has a prescribed failure rate for which data are contained in the literature. These are not reproduced here. Instead, a simple logic diagram showing how these components interrelate and how their failure might affect the process, is appended. This diagram shows the failure rates presumed and the reference list attached, indicates sources.

The mooted computer-control system has separate input and output cards and a back-up microprocessor. It controls all actions of all components of the hydrogen plant in accordance with pre-programmed cues. Such cues may be varied if required. If a failure of the main processor occurs, the other takes over ‘seamlessly’ so that even the operators in the plant control room would not initially realise that the change had occurred.

If the computer system fails completely, consequent on power failure or inherent failure, first, the Uninterruptible Power Supply *(UPS)* would take over and continue the operation of the system, for a period of some 12 hours. Secondly, all controlled items would 'fail' in safe mode that is, an Emergency Shut Down *(ESD)* would occur. Valves for example, would fail open or closed as required for safety. An emergency shutdown can also be caused by operating personnel, either via control room software or by buttons which are 'hard-wired' to cut power to the computer and force various items to fail safe.

Again for simplicity, a failure of each and every physical component in the system, as well as failure of the distributed control system is assumed herein and its likelihood established.

Physical component failure rates range from < $1/10^6$ *hrs, to* $9690/10^6$ *hrs*, the latter being for motor-driven compressors. When each of the pertinent failure rates are combined *(see attached logic diagram),* the resulting system failure rate for all mechanical components is $2.23/10^{88}$*hrs* of operation. Unless all components failed into their unsafe condition simultaneously, it would not be possible for high-pressure hydrogen to reach the plant.

Derivation of such a low failure rate is a consequence of the fact that most of the components must fail simultaneously that is, in Boolean algebraic terms the individual failure rates must be multiplied together at *'and gates'*. Where failure of one component or another is pertinent, values are added together at *'or gates'*. As all listed failure rates even the very high ones are fractions, multiplying fractions together rapidly yields the inordinately low failure rate number determined.

Similar logic has been used for the computer control system but simplified and as already indicated, with less reliable data. It is established that the computer will fail completely, $7.5/10^{11}$ *hrs*. Computer failure or power failure results in all components failing safe. Therefore this failure rate need neither be added-to nor multiplied-by the established mechanical equipment system failure rate. Therefore although undesirable, failure of the computer and of all operator interventions which would usually follow such a failure, would cause a safe condition to be arrived at provided that each and every item of plant responded as it is designed to do. Valves for example, are designed to fail closed or to fail open by spring pressure.

Assuming *8000 hours per year* of operation, the 'top event' failure rate becomes *1.78 failures/*10^{84} *years*. More commonly expressed, this is a failure rate of

$$\mathbf{1.78 \times 10^{-84}/year}$$

Not only is an individual risk of 1.78×10^{-84}*/person/year* acceptable in the community but essentially this is saying that the event is close to being impossible.

8 OTHER POTENTIAL FAILURE MODES

8.1 Storage vessel relief

It is mooted that storage of hydrogen is within pressurised road trailers. Each typical trailer contains 12 horizontal cylinders and each cylinder is fitted with a relief device. The principal and purported sole mechanism of failure, is operation of the relief devices which yields an intense vertical jet of flame causing little harm. Each trailer holds *374kg* of hydrogen compressed to *165bar (16,500kPa).* Two or three such trailers are generally present at a facility either to supply their contents or to be filled.

8.2 Hydrogen embrittlement

This causes drastically reduced strength and ductility of steel and also of other metals. In one example, a *1.85m dia x 6.75m* long pressure vessel with *5mm* wall thickness in operation for 12 years, at a pressure of *10,000kPa (10MPa)* burst causing damage estimated at $4 million. Material of construction was *ASTM A517-F* steel. This effect has not been considered in this study as the assumption has been made that materials have been appropriately selected. Some photographs on this topic are appended.

8.3 Use of nickel catalyst - propensity for internal reaction

In some hydrogen-using plants namely oil hydrogenation plants, the combination of spent nickel catalyst and oil is known to be pyrophoric in contact with air. In the reactor, air is absent and therefore this is unlikely to present a hazard. Other aspects of potential for adverse or runaway reaction have not been considered here. Use of low-boiling solvents in hydrogenation should be avoided to reduce the risk of ignition when catalyst is added *(Bretherick Vol 2, 1999).* A proprietary form of Raney nickel catalyst in which the finely divided metal particles are coated with a fatty amine is claimed to be free of pyrophoric hazards if it dries out *(Bretherick Vol 2, 1999*).

8.4 Inappropriate pressure relief system design

Pressure vessels should contain pressure relief devices which are either rupture discs *(RD)* or relief valves, not both. Often, the relief device is a pressure relief valve *(PRV)* with a rupture disc placed below it to prevent clogging of the PRV in normal operation. To compensate for the possibility of prejudicing rupture disc burst pressure by development of pressure between RD and PRV, a pressure switch and pressure alarm are often present. Where there is a tendency for clogging of relief valves, notwithstanding the concerns of environmental agencies, bursting discs should be used, not relief valves.

In hydrogenation plants, it is understood that there is a considerable tendency for reactants to cause blockage and gumming of relief valve components and that this is why rupture discs are present. Typical rupture discs are a composite of stainless steel and teflon, selected to fail at a pressure significantly below that of the relief valve. Provided that the interspace pressure sensor works, any condition resulting in passage of pressure beyond the rupture disc will lead to a ‘latched’ shutdown. Thus pinhole leakage and an excursion above rupture disc fail pressure, will both trigger the emergency shutdown. The latching process adds a further level of safety in that latched events are usually separately recorded and unlatching cannot be enacted without authorisation. Unlatching in this event, would be preceded by replacement of the rupture disc, cleaning, inspecting and testing of the relief valve and cleaning of the interconnecting pipeline.

Where there are sound arguments in favour of rupture disc/relief valve combinations, the solution may be to install a second relief route comprising a rupture disc only, but selected to fail at a pressure above the relief valve pressure, re-rating the relief valve at the same time as required.

8.5 Flame impingement

Flame temperature at the stoichiometric fuel/air ratio *(31.6% H_2 by vol)* is a maximum of *2403ºK (2130ºC).* This is very high and means that it is particularly important to ascertain all potential sources of emission and to design a plant so that flame impingement upon critical components is not possible *(CRC 1977).* The oxy-hydrogen flame is used for high temperature cutting and welding of very thick sections of metal. A jet leak of hydrogen would burn less fiercely than a stoichiometric mixture of hydrogen and oxygen but ramifications could still be serious.

A figure for reactor failure rate by this route has not been found. Prevention would depend on control of leaks, a high level of ventilation and the absolute absence of ignition sources. The former is difficult in view of the small size of the hydrogen molecule.

Flame impingement from a flange leak onto a pipeline was alleged to be responsible for escalation of the 1974 UK Flixborough explosion. Some are of the view that this was the initiating event but others *(Gugan, Bond, Alexander, Thomas*) do not support this view.

8.6 Computer viruses

A recent instance of production plant forced shutdown as a result of a computer virus, suggests the need to consider this possibility in the present context. It is assumed that in plants which are controlled by computer, the computer is a 'stand-alone' system and is therefore not readily liable to virus attack. However, the Holden Adelaide plant which was the subject of the above-mentioned failure, is also understood to have been a stand-alone system.

If a virus caused the subject plant control system to fail this would simply lead on to an Emergency Shut Down (ESD) as indicated elsewhere herein. Although this would have financial ramifications it would not prejudice plant safety.

9 OTHER CONSIDERATIONS

9.1 Societal risk

This is established by assessing the actual population inside the effect distances previously established. In regulatory compliance terms, only societal *(or group)* risk of death is considered and as none of the modelling herein performed indicates deaths off-site, no attempt has been made to establish societal risk.

9.2 Wind direction effects

It is common in modelling like that performed herein, to establish the wind rose for the location and to consider the predominant direction in which an effect travels and then to calculate individual risks for communities in several compass directions around the source. However, this has not been performed as an explosion produces blast effects which are essentially equal in all directions.

10 MORE ON THE PROPERTIES OF HYDROGEN

Ignition temperature is *847ºK (574ºC).* This is high and for example, means that a lighted cigarette or cigar will not ignite it. Smoking should not be permitted however, as the energy in a lighted match or lighter is sufficient to cause ignition.

Flame emissivity is *0.10*. This is low and means that burning is unlikely unless the flame impinges on the person. The low emissivity however, means that the flame is invisible in normal lighting conditions and therefore an observer may be in the flame without knowing it until burning has occurred. The latter effect is enhanced further as hydrogen is also odourless.

Hydrogen is an æsphyxiant gas. Together with absence of odour this means that a considerable æsphyxiation hazard is present in any *confined* plant where hydrogen is in use. A multi-point, cyclic ambient monitoring system analysing oxygen level should be installed in such a plant.

The hydrogen molecule is so small that prevention of all leakage is almost impossible.

Flange leakage can be more properly managed than taping and periodic use of a hand-held analyser, by the banding of flanges and passage of a designed outlet tube to a safe place.

The extreme lightness of the hydrogen molecule means that in unconfined situations such as transfer pipelines, explosion is unlikely consequent upon propensity for leakage to disperse upwards. Although the ignition energy is extremely low, detonation energy is high. The converse of this is the case as soon as any form of *confinement* is present. Even the 'confinement' of an open plant structure has been shown to increase the likelihood of explosion *(Flixborough Disaster, UK, 1974; CRC 1977).*

Sudden release of hydrogen at *7900kPa (79bar)* or greater can cause spontaneous ignition as a result of the reverse Joule-Thompson effect. Presence of fine particulate metal or rust can increase likelihood of this *(CRC 1977*).

Occurrence of flash fires and explosions on discharge of the hot products of catalytic hydrogenation of vegetable oils have been attributed to presence of *phosphine* from the *phosphatides* present eg in rapeseed and linseed oils *(CRC 1977).*

CONCLUSIONS

- If a confined or semi-confined hydrogen explosion occurred at the facility of the scale considered herein, sufficient injury and property damage would be caused on and off site, to cause authorities and the courts to take actions which might lead to its closure

- *40kg* of hydrogen leakage is somewhat unrealistic as practices indicate that this could only happen if numerous safeguards were breached or failed

- A more realistic but also unlikely event is the emission of *3.9kg* of hydrogen, leading to explosion. This too would cause some off site effects and on site, anguish and possible loss of life

- It is likely that an actual hydrogen escape would rise away from the facility so quickly that an explosion if it occurred, would be in the air above the plant rather than within it, provided that its design allows a sufficiently open structure to permit this. Consequently it is considered that the extreme buoyancy of hydrogen would mitigate strongly against an event being as significant as indicated herein

- The individual risk in the surrounding community using raw pressure vessel failure data, is below the criterion informally used for the control of major hazard facilities by the authorities in Australia

- The quantity of hydrogen gas used or generated at such facilities is sufficient to warrant their definition as Major Hazard Facilities *(MHF)* under the *Occupational Health and Safety (Major Hazard Facilities) Regulations 2000*

- Individual risk criteria and likelihood of explosion occurring at all, are so low that no personnel present on such a site face an unacceptable level of risk

- The actual explosion risk level calculated for the whole system comprising the hypothetical hydrogen-using plant is such as to indicate that an explosion of the scale considered, is essentially impossible

- The all important conducting of *HAZOP* studies in instances like this must involve all of the right personnel and free flow of information without embarrassment and without holding back. Without this, the process however rigourous, would fail

11 FOOTNOTE

Appendices are omitted herein due to space limitations but will be presented at the conference and on request. They comprise :-

- *A series of four cartoons highlighting the importance of applying common sense as well as mathematics to risk assessment*

- *A notional risk assessment worksheet where risk is depicted in the range 1 to 25 and serious risks are 10 and over*

- *Detailed list of references cited (some are included in the text)*

- *Maps of human injury radii*

- *Maps of property damage radii*

- *Process flow diagram for the hypothetical plant*

- *Control system logic diagram for the hypothetical plant*

- *Hydrogen brittle fracture and vessel failure photographs*

A COMBINED OPERATIONS AND MAINTENANCE STRATEGY FOR IMPROVED PLANT PERFORMANCE: SURVEY RESULTS ANALYSIS

Fernando MR and Cheong F

School of Business IT, RMIT University, Australia

Abstract: We present, discuss and analyse the findings of a recently completed questionnaire survey conducted within a large Australian organization in relation to a case study based empirical research project on above topic. Key aspects of the topic, including its research trends, rationale, barriers, influencing factors and performance measures are discussed. The results, representing perceived views of a population sample consisting of maintenance and production staff on relevant aspects, are analysed and discussed. Overall results were encouraging with a majority of participants in favour of the concept. There was also significant opposition or disagreement, challenging potential implementers to properly address the needs of this opposing minority and gain their support prior to the introduction of such strategies.

Key Words: Maintenance Strategy, Plant Performance, Empirical, Performance Measures.

1 INTRODUCTION

Having invested heavily in the state-of-the-art, automated plant and machinery, modern manufacturers have managed to reduce their production or processing costs significantly with improved delivery and customer service performance. However, there remain future challenges, especially in terms of initiating continuous improvements to further improve their plant performance; a challenge that confronts plant managers in an ever increasing competitive pressures due to globalization [1, 2].

Traditionally, the function of production has been considered as primarily value adding and directly responsible for processing of product or service. However, with the introduction of high technology, automated plant, maintenance is now becoming a key driver of plant performance, due to its contribution through expert technical and reliability support. While bad maintenance leads to unreliable plant, outcomes due to poor operational practices could be drastic [1, 2]. An identifiable dynamic interaction [3] exists between production and maintenance, causing adverse effects on overall plant performance, more often than not. Hence, the challenge is to eliminate such negative outcomes through a common focus on overall plant performance. This is achievable by understanding the value adding role of each other, towards the overall plant profitability and competitiveness.

This dynamic interaction has prompted research attempts to combine maintenance with production and other manufacturing disciplines [3, 4]. Encouraged by demonstrated positive outcomes of such initiatives, we proposed a Combined Operations and Maintenance Strategy (COMS) [5], as one solution aimed at improved plant performance through collaboration teamwork between production and maintenance, by focusing on common business objectives. A well-defined COMS, jointly formulated through an interactive team process, could provide modern manufacturers with a distinctive competitive edge. Furthermore, initiatives similar to COMS could play an important role in modern, automated plant, where maintenance play a dominant role in providing timely solutions to enhance uptime and plant reliability through highly skilled technical staff. Therefore, it is to the benefit of both production and maintenance managers, that they exploit the potential benefits of COMS and its effective implementation. Hence, we believe an empirical study on COMS is justified, to serve the above purpose.

2 SURVEY DETAILS

Our study involved a recently completed questionnaire survey targeted at an audience comprising of maintenance and production staff of a large Australian organization, seeking their perceived views on COMS implementation. The organisation provides collection, distribution and delivery services of various articles across Australia, serving a wide range of domestic and international clients. This survey was conducted within seven of their processing plants situated in major cities of Sydney (SYD), Melbourne (MEL), Adelaide (ADE), Brisbane (BRI-1 & BRI-2), Perth (PER) and Canberra (CAN).

The organisation employs nearly 23000 people with annual revenue of $ 4.3 billion and $ 525 million before tax profit reported during 2004-05. With a total asset base of $ 3.8 billion, the organization has invested in modern, the state-of-the-art processing plants in each state, and has a highly sophisticated road and air transport network. A population sample from both maintenance and production staff participated in the survey, except in SYD and BRI-2, where timely approval could not be obtained for the involvement of production staff. A total of 148 staff members from maintenance (Maint) and 116 staff members from production (Prod) participated the survey, with a targeted audience comprising of technicians, team leaders, technical managers, production managers, process leaders and production operators.

We visited all sites and personally invited staff to participate in the survey, except for CAN, where the survey was carried out by respective managers according to guidelines. Since the focus of study is maintenance management, we invited all staff from maintenance function to participate, except for those who were on leave. Due to very large production staff population, only a representative sample of production function was invited to participate. A breakdown of all participants is shown in Table 1 below. Response rate of maintenance staff is shown in Figure 1.

Table 1

A Breakdown of all Participants

	MLB		ADE		BRI-1		CAN		PER		SYD		BRI-2		Overall	
	Prod.	Maint	Prod.	Maint	Prod.	Maint	Prod.	Maint	Prod.	Maint	Prod.	Maint	Prod.	Maint	Prod.	Maint
Manager	3	1	2	0	2	1	2	1	3	1	0	1	0	0	12	5
Team Leader	13	9	16	3	20	4	11	5	9	0	0	4	0	2	69	27
Staff Member	5	42	13	14	2	14	3	5	12	11	0	25	0	5	35	116
Total	21	52	31	17	24	19	16	11	24	12	0	3	0	7	116	148

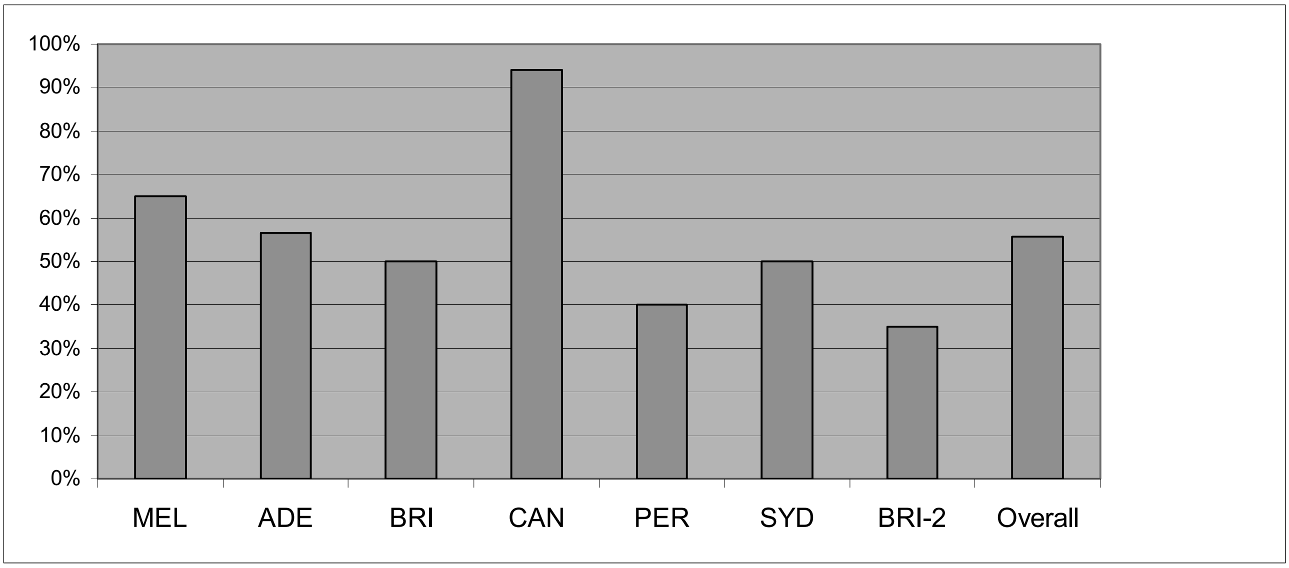


Figure 1.Response Rate of Maintenance Staff

3 RELATIONSHIP BETWEEN PRODUCTION AND MAINTENANCE

Poor maintenance causes disruptions and unavailability of plant. Lack of maintenance planning affects production performance through redundant manpower and loss of output, eventually declining the overall business performance [2]. It is also evident that poor production practices, lack of production planning and basic operator care, often cause reliability problems and downtime. Although maintenance has been historically blamed for equipment down time and for higher costs, current research reveals that a majority of maintenance costs are driven by bad design, poor production practices and process controls [6, 7].

The above dynamic interaction between maintenance and production necessitates close liaison and coordination between the two functions, because they can no longer be treated in isolation. Instead, the two functions should complement their mutual efforts to achieve overall business objectives. Eventually, both functions have to realign their individual strategies based on common priorities, including a most important one in continuous improvement of plant performance [5].

In our survey, we sought the opinion of staff on the relationship or interaction between production and maintenance. They had to choose between two options; (a) *Independent* of each other, each contributing separately towards plant performance, and, (b) *Interdependent & Interacting* with each other to influence plant performance. A comparison of their selections is illustrated in Figure 2.

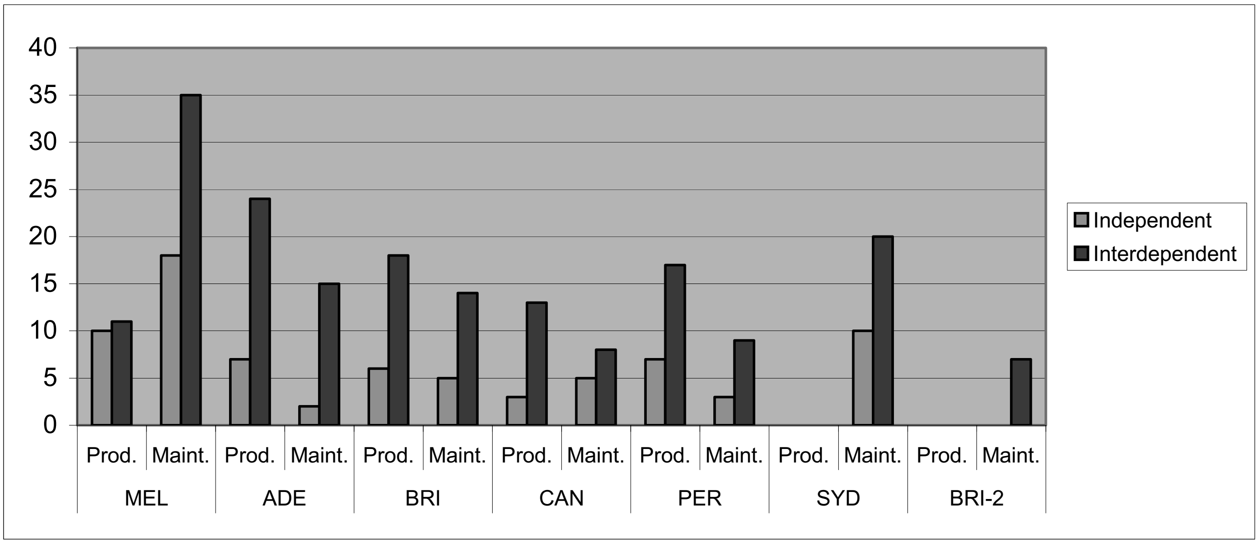


Figure 2.Relationship between Production and Maintenance

The results indicate that a majority of staff from both production (Prod) and maintenance (Maint) consider these two functions as *interdependent and interacting* with each other to influence plant performance.

In terms of the contribution of each function towards influencing overall plant performance, we invited staff to rate each function, between a scale of 1 to 5, 1 representing a very high, and 5 being a very low. The results are illustrated in Figure 3.

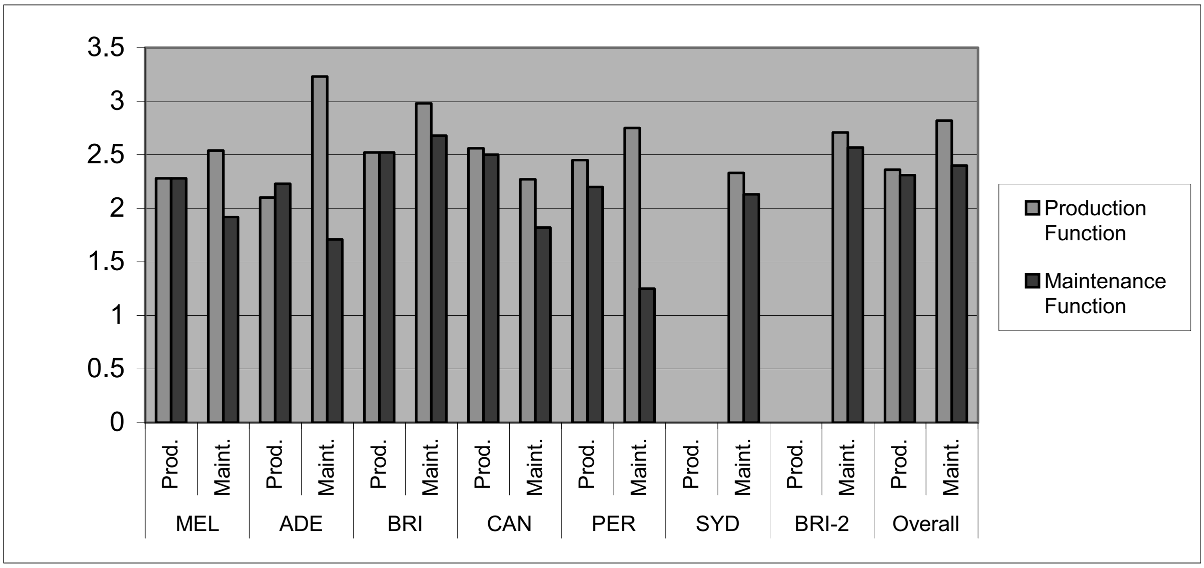


Figure 3. Contribution or Level of Influence over Plant Performance

It is evident from the results, except for a slight variation from production staff in Adelaide, both production and maintenance staff at all plants placed maintenance at a higher rating than production. This confirms that they believe maintenance has a greater influence over plant performance. In general, ratings of both functions are much closer to each other, suggesting that staff considered both functions playing equally important roles.

4 RESEARCH RATIONALE

Maintenance management has emerged from being considered a marginal activity or an unavoidable cost, to that of a strategic function which provides cost effective solutions to achieve overall business objectives. Many researchers have identified the need to combine maintenance with other manufacturing disciplines such as production. A majority of these research initiatives have demonstrated positive business outcomes [3, 4, 8-11].

Although production is the primary value adding plant function, there is enough evidence to show that maintenance adds value to production process by providing an essential support service to ensure plant quality and reliability. Highly dominant role played by skilled technical staff in maintaining high technology plant in many sites demand maintenance staff to be equally competent in handling all production situations with confidence. Similarly, production staff has to be fully aware of the nature and complexity of high technology plant, in order operate the plant effectively and confidently. A well-defined COMS, involving combined team efforts between the two functions, will help both maintenance and production functions to achieve these goals, leading the plant to high profitability and competitiveness.

The following extracts from current research also justify the need for COMS or similar initiatives.

- Resource utilisation is critically linked with plant availability, which again depends upon maintenance reliability. Hence, a sound maintenance policy maximises profits [3].
- Competitive pressures demand strict cost control through reassessment of maintenance practices, where technical staff should gain high production process competency [10].
- An approach to link performance with a combined maintenance and production strategy would supplement the existing theory and practice of operations strategies [2]
- The involvement of all disciplines, is an essential feature towards achieving overall production efficiency [12].

In the survey, we invited staff to select from the following three options, the best choice that they think, would improve plant performance:

Option A Separate strategy for each department, in line with their functional objectives.

Option B Separate strategies formulated by each department, but in consultation with one another, in order to complement the business objectives of each function.

Option C A combined operations and maintenance strategy (COMS), jointly created, through an interactive team process, in line with the overall business strategy.

The level of responses received by each option for this enquiry is shown in Figures 4 and 5.

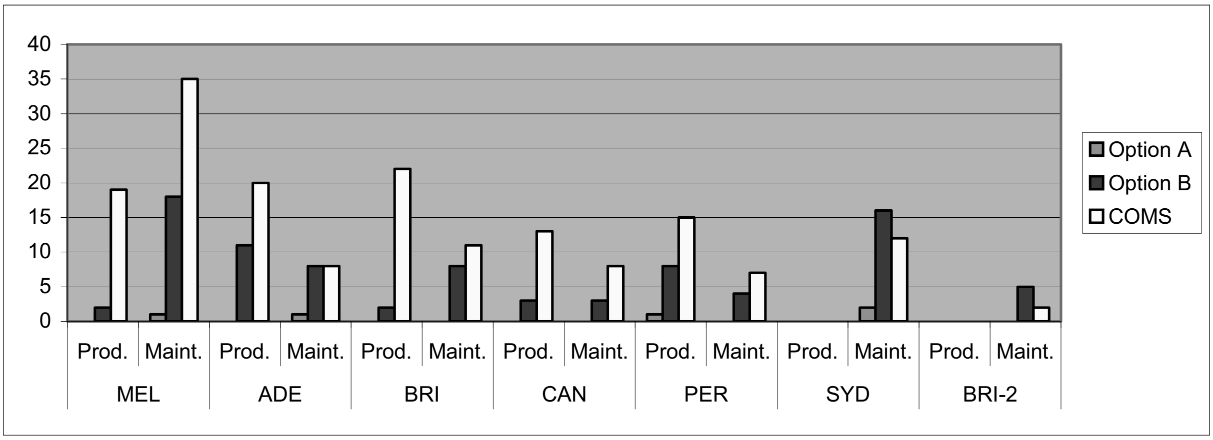


Figure 4.Best Strategic Option, Perceived by Staff from Each Site

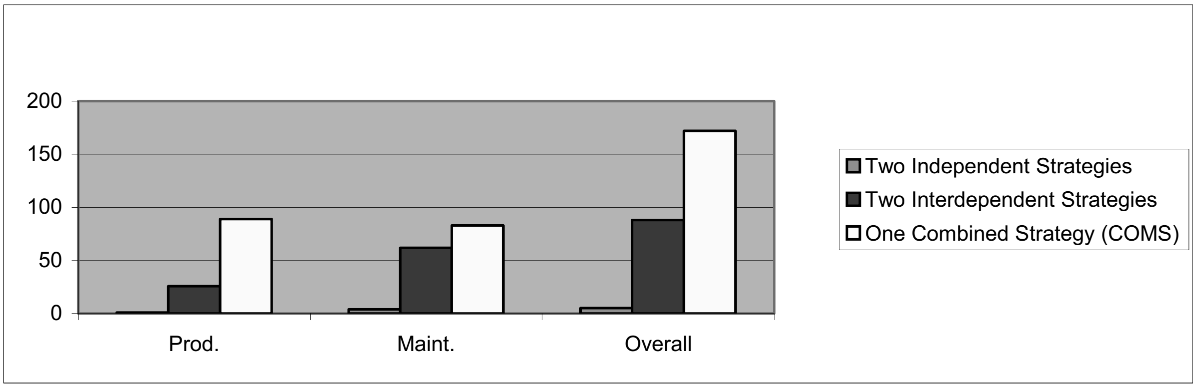


Figure 5.Best Strategic Option, Perceived by Overall Staff

The above results demonstrate a strong support for COMS among production staff at all sites. However, the opinion of maintenance staff varied between sites. While a majority of maintenance staff at MLB, BRI, PER and CAN preferred COMS, a majority of maintenance staff at SYD opted for two interdependent strategies (option B). ADE remained neutral.

Although a majority (80) of maintenance staff was in favour of COMS, a significant number (60) preferred option B. This could mean that any future COMS implementation attempts may face with resistance or reluctance to participate, especially from this section of maintenance staff. Technicians at MLB and CNB will possibly offer more support and lesser objections, and they would make ideal sites for future pilot programs or field experiments on COMS.

The results relating to the current strategic approach taken by management, as perceived by staff, are shown in Figure 6.

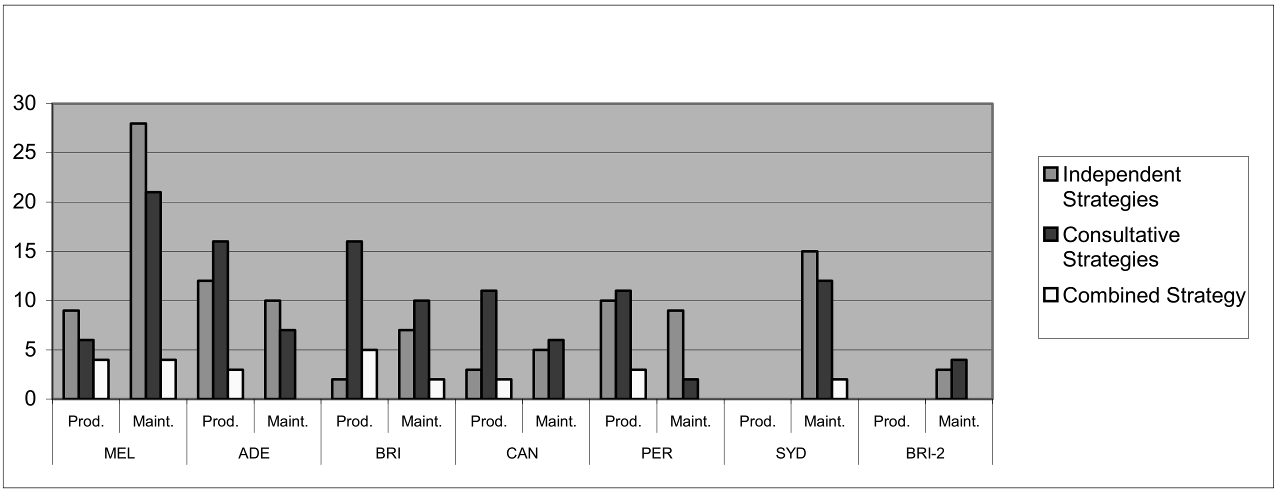


Figure 6.Current Management Strategy, Perceived by Staff

Many believed that current approach is consultative. Given the choice to decide on their own strategy, the staff selected the given strategic options as illustrated in Figures 7 and 8.

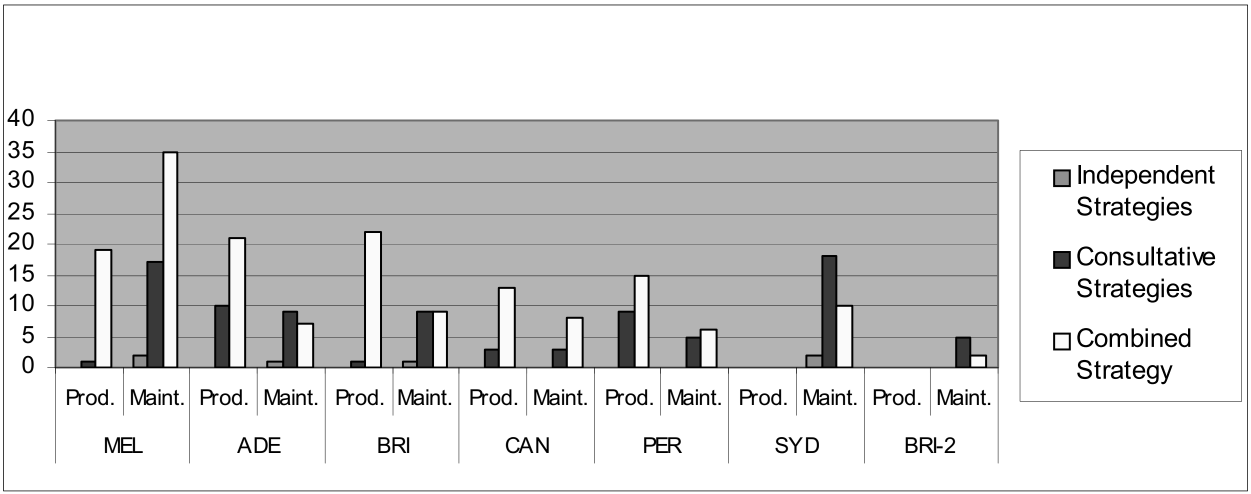


Figure 7.Strategic Approach chosen by Site Staff

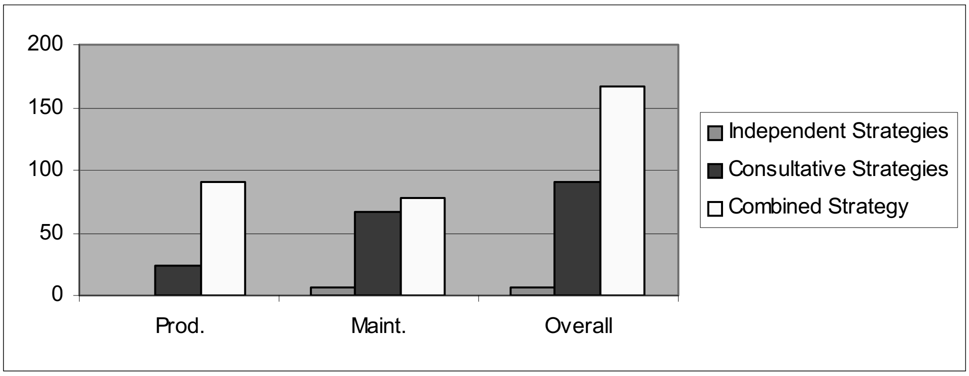


Figure 8.Strategic Approach chosen by Overall Staff

While production staff preferred COMS, maintenance staff was equally divided between consultative and combined approaches. When asked, whether they consider COMS, based on common business objectives, as a genuine industry need for their current plant environment, the staff responded in a way as shown in Figures 9 and 10.

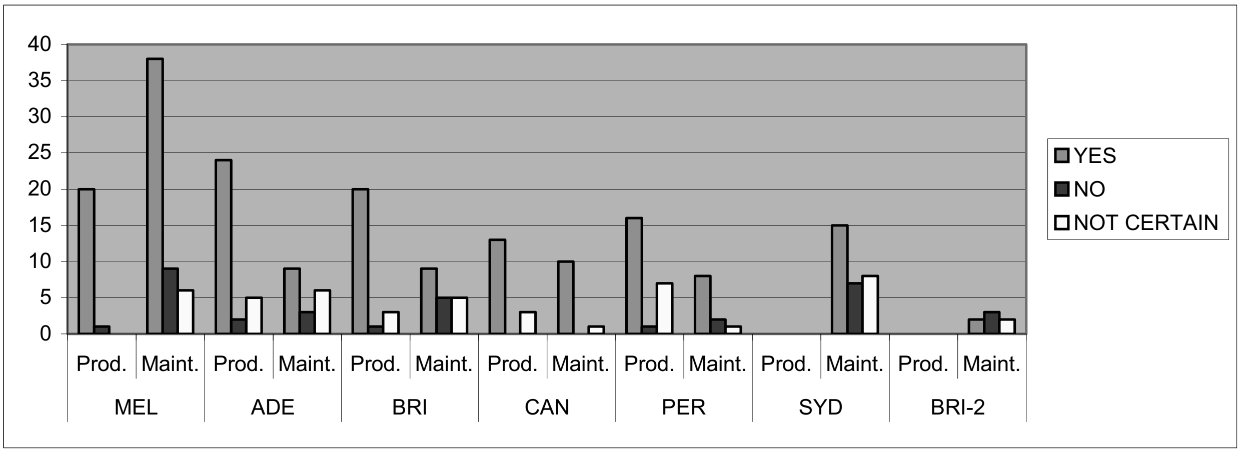


Figure 9: Is COMS a genuine Industry need? - Site Staff

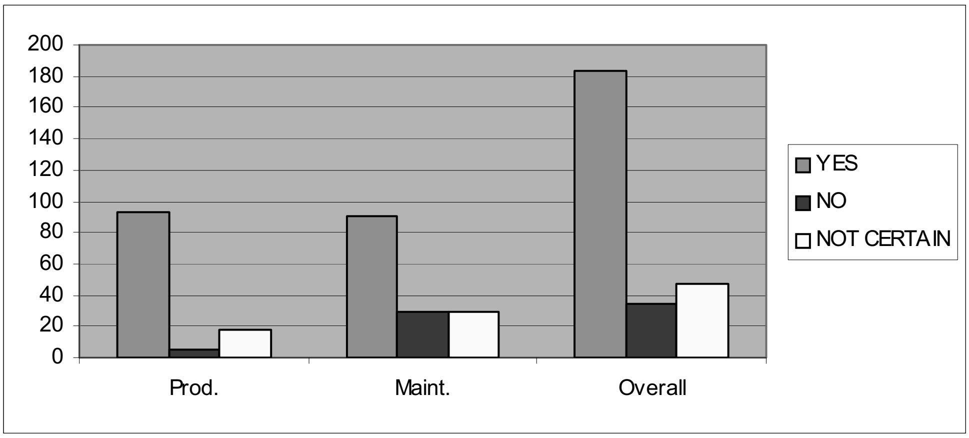


Figure 10.Is COMS a genuine Industry need? - Overall Staff

The above figures indicate that a clear majority consider COMS as a current industry need.

5 BARRIERS OF COMS

Research identifies a historically existing conflict of interest between operations and maintenance, due to lack of communication, and other reasons. A successful COMS could only be achieved through total elimination of such internal barriers through strategic change. However, implementing strategic change involves risks. Risks often discourage decision makers. Sutton [13] claimed that 70% of change project initiatives in the UK failed due to lack of planning, shared vision and team involvement. Other barriers confronting COMS are natural human fear and resistance to change due to uncertainty and the loss of power and authority [14].

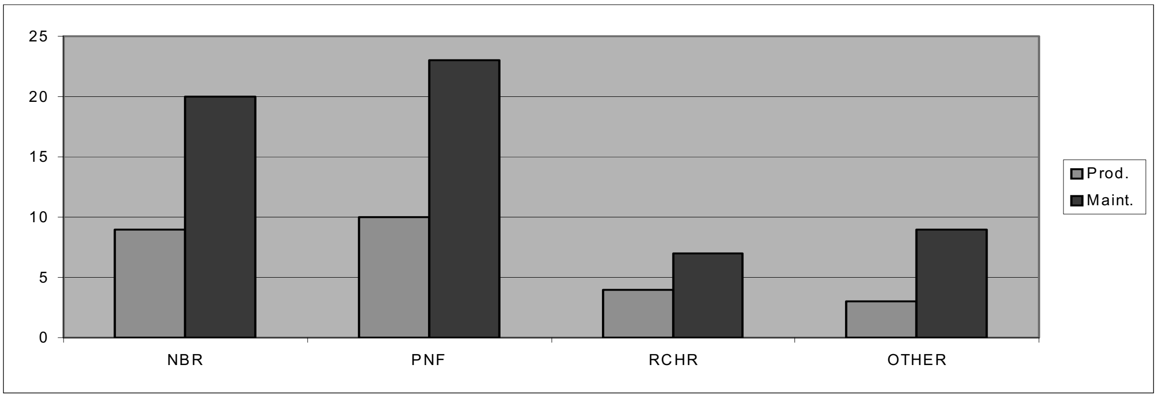
From those who preferred two interdependent strategies, but not COMS, we asked them to select their possible reason for not selecting COMS, from four given options:

Option 1 COMS will not bring any better results than two interdependent strategies (NBR)

Option 2 COMS is practically not feasible for their plant (PNF)

Option 3 COMS is too radical and the risks involved are too high (RCHR)

Option 4 Others, please specify (OTHER)



The number of times each option was selected by the staff is shown in Figure 11.

Figure 11.Reasons for not selecting COMS

The above results indicate many believed that COMS is not practically feasible for their site, or they did not believe that COMS will bring in any better results than two interdependent strategies.

From those who preferred options A & Option B in section 3 above, we asked them to identify major barriers of combined or collaborative strategies, from four given options as follows:

Option 1 Fear and resistance to change by management and staff (FRC)

Option 2 Opposition from those who may lose their power and authority (OPP)

Option 3 Reluctant or failure by plant management to identify such a need (RFM)

Option 4 Lack of communication and understanding between functions (LCU)

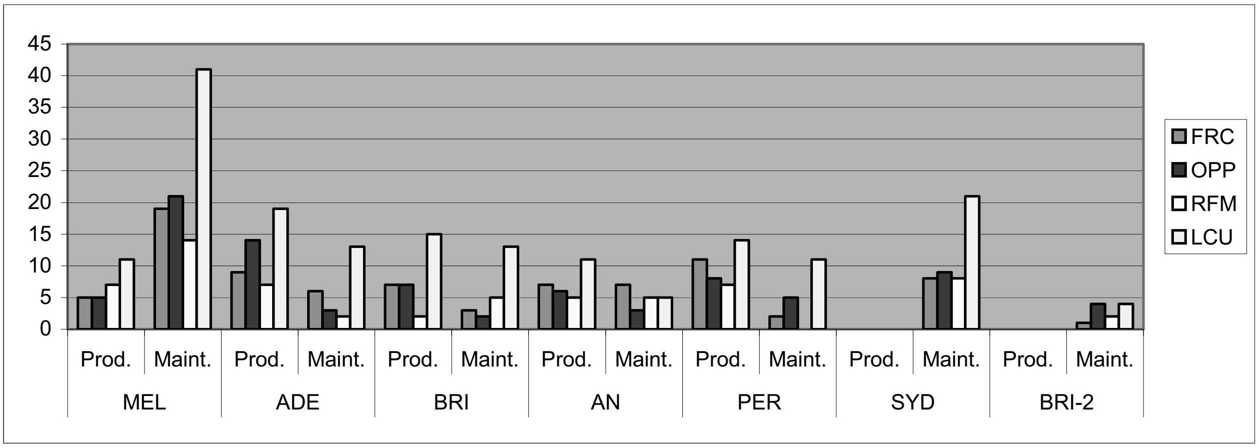


Figure 12.Barriers confronting COMS, Site Wise

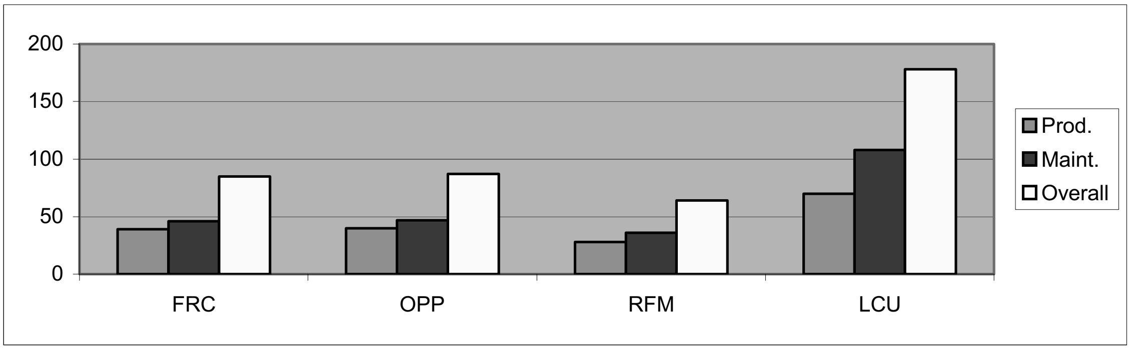


Figure 13.Barriers confronting COMS, Overall

Total number of times each option was selected, are shown in Figures 12 and 13 above, and the results confirmed our earlier research finding that lack of communication is a major barrier of COMS, and proves that improved communication and better understanding is vital to make COMS a successful initiative.

6 PROMOTING COMS

Communication gap between the two functions can be reduced by understanding and focusing on common business goals and by having a shared vision. Once production and maintenance clearly identify each other's strategic role in adding value to the overall business, their mutual efforts could be supplemented. Multidisciplinary teams or autonomous groups [12] may help to have better understanding of mutual needs and lead to improved harmony. Education and training also play a vital role in bridging such gaps through regular meetings and focused groups, fostering collaboration.

A remarkable lack of trained technical personnel to perform primary maintenance activities has been lately identified. This gap will have a greater impact on the availability of researchers involved in maintenance related areas [15]. By promoting industry based maintenance research through sponsorships or grants can encourage maintenance professionals into active research. Participation of academics and experts in industry-based research could also help.

In our survey, we asked those who preferred combined or collaborative strategies to rate the level of influence of the following action items, towards promoting such strategies.

Option 1	Initiate open dialog between the two functions (IOD)
Option 2	Involve operators and Technicians in decision making (IDM)
Option 3	Promote and create cross-functional teams (PXFT)
Option 4	Educate plant management on COMS (EPM)
Option 5	Establish sound theoretical knowledge of COMS (EST)

Weighted average rating of each action item, as perceived by staff, is shown in Figure 14, with 1 representing a very high and 5 representing a very low.

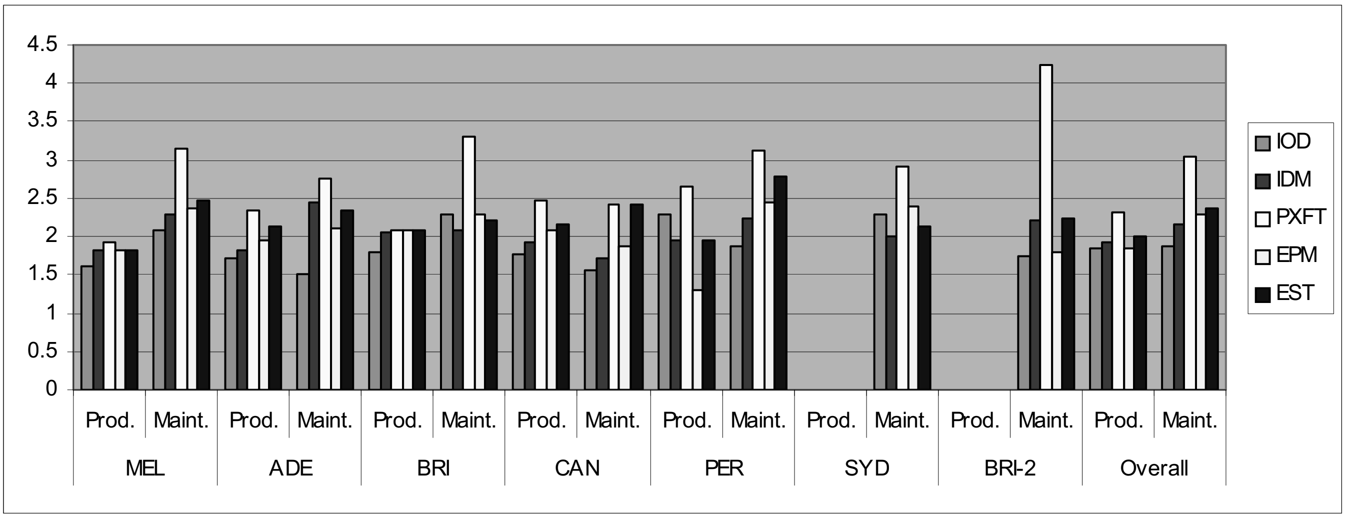


Figure 14.Factors influencing Combined Strategies

In all sites, both production and maintenance staff rated IOD (initiating open dialog) as the most influencing action item. Production staff also found EPM (educating plant management) equally important. Staff from both sides identified the need to initiate open dialog between them, possibly reassuring us of an existing lack of understanding between the two functions. With the exception of PXFT (promoting cross-functional teams), all staff in general found all action items of having a more than average influence in promoting COMS. It is also evident that maintenance staff do not favour PXFT, clearly giving us an indication that technicians prefer to have their own identity.

7 MEASURING PERFORMANCE UNDER COMS

Success of any business entity depends upon its ability to meet the desired level of performance. Hence, it is important that we establish suitable indicators of plant performance, and use those yardsticks to measure our business success. There are a few performance measures we have identified as more appropriate under COMS like initiatives.

We found Overall Equipment Effectiveness (OEE) as the most relevant performance indicator for COMS. OEE is easy to understand, flexible, can be used by operations and maintenance, and OEE tracks progress on equipment improvement on regular basis. The strength of OEE is its ability to represent the combined efforts of both operations and maintenance in one measure. OEE is defined by the following equation [12]:

OEE = Equipment **Availability** x **Performance** Efficiency x **Quality** Rate

Whereas;	Availability	= Uptime / Scheduled Time	Industry Benchmark = 90%
	Efficiency	= Actual Output / Scheduled Output	Industry Benchmark = 95%
	Quality Rate	= Production less defects / Total production	Industry Benchmark = 99%

Therefore, Benchmark of OEE = 90% x 95% x 99% = 85% [16].

One drawback of OEE however, is its focus on equipment types, but not on the plant as a whole. Therefore, OEE can only be applied to plants where equipment can be separately identified with clearly established boundaries. OEE has no direct financial interpretations, and therefore, may not attract financial managers and investors.

An alternate solution to OEE, is the Balanced Scorecard (BSC). BSC is an indication of four strategic perspectives of performance, namely, financial, customer, internal processes, and learning and growth. BSC translates strategy into more tangible and actionable objectives, related measures, targets and action plans. BSC could focus on key maintenance related factors, which are critical towards contributing to business success. BSC, while being capable of providing a holistic assessment representing a range of critical issues, is complex and calls for a higher level of understanding. While OEE may be suitable at functional level of the plant, BSC is more suitable at strategic business level [17].

On scale of 1 to 5, with 1 indicating a very high and 5 indicating a very low, we invited staff to rate most relevant and realistic performance measure from four given options, namely OEE, BSC, Unit Cost (UC) and Return on Investment (ROI). The weighted average ratings of each measure, as per results are shown Figure 15.

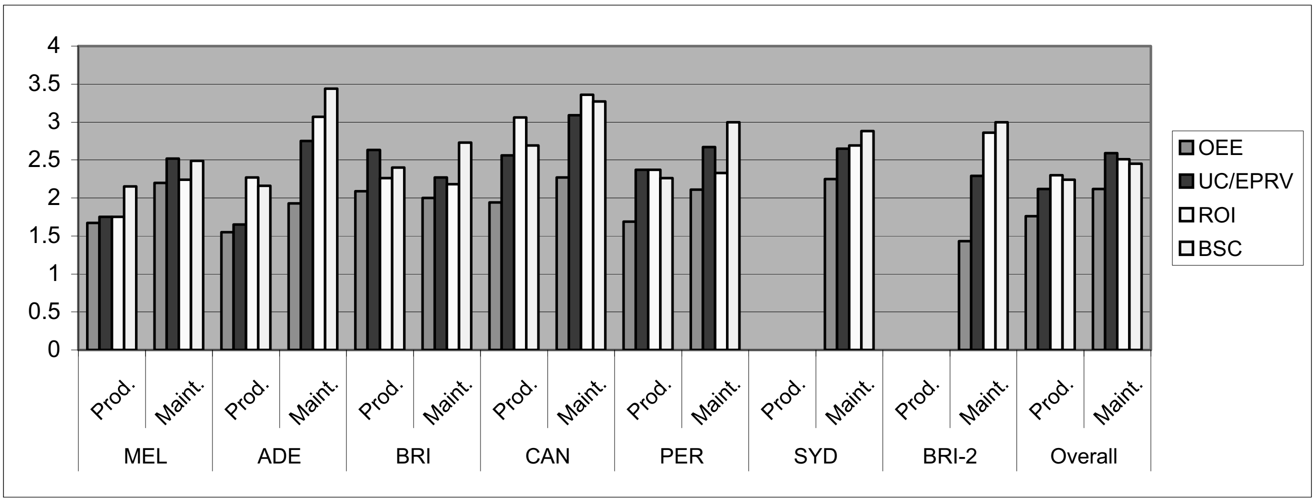


Figure 15: Key Performance Indicators, Weighted Average

From the results, it is evident that OEE is the most preferred option, naturally due to its awareness among the staff within all sites. BSC was not preferred by both production and maintenance staff. Rating of ROI as the second best choice indicates a strong bias towards ROI among all the staff. This appreciation of business returns by the staff is encouraging and the introduction of a monetary based KPI, could motivate the staff towards optimum results.

8 CRITICAL FACTORS OF COMS

There are many concepts and philosophies associated with maintenance management. Most frameworks consist of a critical set of strategic elements, which are needed to be addressed in an appropriate manner to steer the process towards desired goals. These elements could be critical action areas [12] or critical variables which control the effectiveness of the process [4]. Model based approaches usually call these as variables. The effects of these variables, direct or indirect, along with relationships they hold with each other, determine the effectiveness of our process.

We have come across a myriad of variables relating to COMS. We have carefully selected a few case sensitive variables from the organization of our study for our survey. They were; goals and strategy, human resource aspects, organizational aspects, support mechanisms, and tools and techniques [17]. Our participants were invited to rate the contribution of each component on a scale of 1 to 5, with 1 representing a very high and 5 representing a very low. Weighted average ratings given to each option are shown in Figure 16.

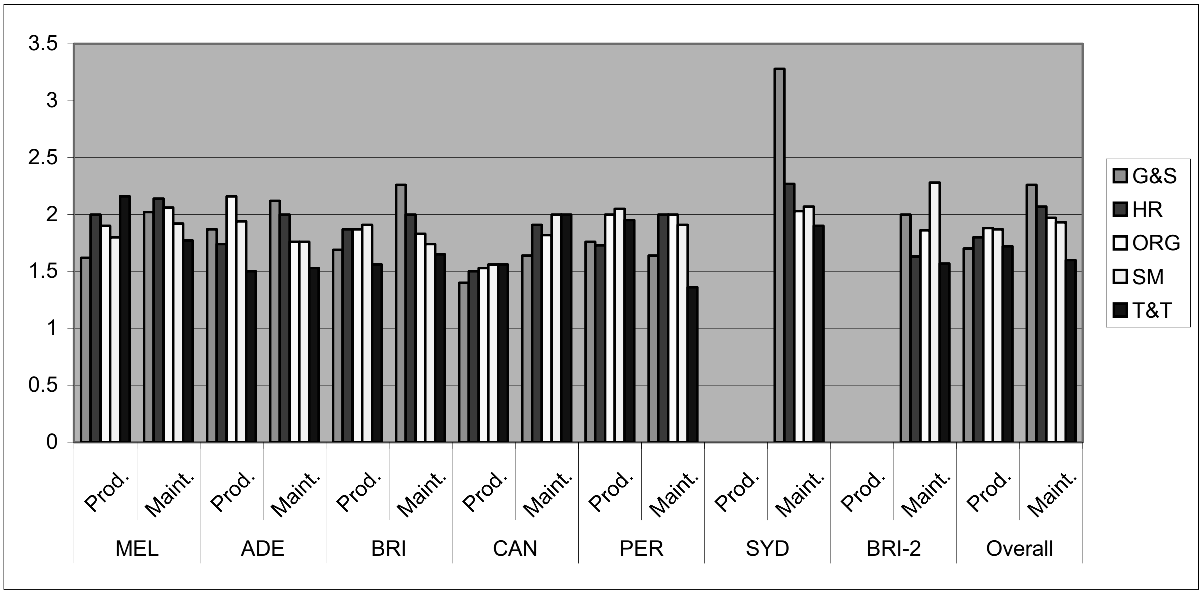


Figure 16. Contribution of Strategic Components

Production staff rated all variables of having equal rating, whereas maintenance considered tools and techniques as having higher influence and goals and strategy with a lower rating.

9 CONCLUSION

Continuous improvement of plant performance is a challenge faced by modern manufacturers striving for profitability and competitiveness. Survey results confirm that maintenance play a dominant role in effectively contributing to this task, especially in these heavily automated high technology plants. Proposed combined strategy aimed at exploiting the true potential of the dynamic interaction between maintenance and operations to improve plant performance, has been accepted by a majority. However, there was a significant number who preferred collaboratively implemented, separate functional strategies. They reasoned that a combined strategy may not bring in any better results than having two collaborative strategies. Feasibility of COMS implementation in these sites was seen by them as impracticable.

Justification of COMS in similar plant environment is strengthened by the acceptance of COMS as a current industry by a vast majority. The lack of communication (and understanding) and fear (and resistance) to change, are two major barriers confronting COMS initiatives. Initiating open dialog between the two functions was seen by many as key influencing factor, overcoming many barriers. OEE was accepted by a majority as an appropriate performance measure, with a significant number demonstrating their appreciation of ROI, also as a key measure.

The overall results of our survey were encouraging, with a clear majority in favour of COMS. However, there are key issues to be addressed before we could gain support from a doubtful, but a significant minority who still prefer two individual strategies. Even though the results seem to verify current research findings on this topic, the final outcome of successful COMS implementation will depend upon the ability of both production and maintenance managers in overcoming the barriers and hence reducing the communication gap.

10 RFERENCES

[1] Moore R., (2004) Reliability Leadership for Manufacturing Excellence: Reliability-based Manufacturing. Vol. V1, Knoxville:TN, The RM Group, Inc.

[2] Jonsson, P. (2000) Towards an holistic understanding of disruptions in Operations Management. Journal of Operations Management,. 701-718.

[3] Vassiliadis, CG et al. (2000) Simultaneous maintenance considerations and production planning in multi-purpose plants*, Annual RELIABILITY and MAINTAINABILITY Symposium*: IEEE.

[4] Vatn, JP, Hokstad, P. & Bodsberg, L. (1996) An overall model for maintenance optimization. *Reliability Engineering & System Safety,* **51**(3), 241-257.

[5] Fernando, MR. & Cheong. F. A combined operations and maintenance strategy for plant performance optimisation. *ICME 2003,* Proceedings of the *9th International conference on manufacturing excellence,*. Melbourne, ICME.

[6] Moore, R. (1999) MAKING COMMON SENSE COMMON PRACTICE: models for manufacturing excellence. Houston; Texas, Cashman Dudley. 356.

[7] Oliverson, R.(1997) Preventable Maintenance Costs More Than Suspected. *Maintenance Technology*

[8] Cassady, DR et al., (2000) Combining preventative maintenance and statistical process control. IIE Transactions 32:,471478.

[9] Davies C &Greenough, RM. (2002) Measuring the Effectiveness of Lean Thinking Activities within Maintenance. *Maintenance Journal*, **15**(3), 8-14.

[10] Hipkin, IB & De Cock, C. (2000) TQM and BPR: lessons for maintenance management. *The International Journal of Management Science,* **28.** 277-292.

[11] Samanta, B, Sarkar, B, & Mukherjee, SK. (2002) Design of an Optimal Maintenance Strategy for Mining Machinery through AHP. *Maintenance Journal,* **15**(2). 14-22.

[12] Nakajima, S. (1988) Introduction to TPM. Cambridge: Productivity Press.

[13] Sutton, B. (2000) QA's Chanj 2000 Game challengesdelegates to face real-life twists and turns off a typical change management project. M2 Presswire; M2 Communications Limited.

[14] Fernando, MR & Cheong, F. (2004) Towards Developing a Model for a Combined Operations and Maintenance Strategy. ICOMS'04 Proceedings of the International Conference of Maintenance Societies 2004, Sydney, MESA.

[15] Bradshaw, L. (2002) Business Success via Better Maintenance. *Maintenance Journa, U.K.* **15**(2), 4-13.

[16] Tsang, AHC. (2002) Strategic dimensions of maintenance management *Journal of Quality in Maintenance Engineering,* **8**: 7-39

[17] Schmidt, S. (2003) Just-in-Time production. New Engineer, 6(3).15-19

GALVANIZED STEEL IN ASSET DESIGN

Peter W. Haberecht[a] and Emmanuel Pimentel[b]

a: Materials and Environments (MATENV) PO Box 18061 Collins Street East Vic 8003

b: Galvanizers Association of Australia Lvl 5 124 Exhibition Street, Melbourne 3000

Abstract: After Fabrication Hot Dip Galvanized Steel is used extensively in infrastructure and large engineering applications. Galvanized steel is increasingly being considered as a "composite material" rather than a component of a duplex system (coating on substrate) and its economic and physical advantages extend beyond its traditional role of corrosion protection. This paper will examine the economic and engineering advantages of using galvanized steel. It will be argued that the capabilities of hot dip galvanized steel should be assessed with all engineering materials, at the design and specification stage and not just as an after thought for asset corrosion protection. An emphasis on a more holistic and specific maintenance approach into the life cycle of assets and their component materials has great economic and efficiency benefits. The example used relates to studies performed on lattice towers found in power distribution.

Key Words: zinc, after fabrication hot dip galvanizing, structural steel, durability, AS/NZS4680, duplex systems, cost benefit analysis

1 INTRODUCTION

The extraction of iron from its ore initiates an irreversible and natural process: its corrosion and tendency to return to its natural state. Unless steel is protected, it will corrode in most environments and compromise the structural stability or usefulness of the fabrications that it composes.

Corrosion of steel and its mitigation, repair and ultimate replacement is estimated to cost Australia between 3 - 5% of GDP (this is similar to studies done in the USA). On current GDP statistics, this equates to a figure of $25 - $41 billion. Thus, corrosion prevention and mitigation are not only an essential element in the economic utilisation of steel, but also in the increasingly scrutinised use of both public and private funds. [1] [2]

The selection of the appropriate corrosion protection for steel can result in initial economic savings coupled with substantial economies in service. These economies can be the result of the reduction or elimination of maintenance and their associated labour costs and lost service time and also in the deferral of the replacement of structures and equipment.

Hot dip galvanized steel is used extensively in most infrastructure applications. Although galvanizing is traditionally regarded as a corrosion protection system for steel, it has other benefits such as improved abrasion resistance and ease of transport (due to the fact that the hardness of the iron-zinc alloys formed during the galvanizing process can be around 50% greater than that of the base mild steel) and greater variability in the conditions under which it can be installed. In fact, the hardness of the iron-zinc alloys formed during the galvanizing process can be around 50% greater than that of the base mild steel.

Galvanizing specifically uses zinc to cover and protect steel from corrosion and it does this via a number of different mechanisms: physical barrier protection, the development of a carbonate patina envelope, and cathodic protection. Although the most easily identifiable examples of galvanized structures include road crash barriers, boat trailers and electricity transmission towers, galvanizing is now used in most structural engineering applications requiring corrosion protection or abrasion and impact resistance.

One of the methods of protecting a steel substrate is the application of a paint coating. Paints have a durability that ranges from a few years to a couple of decades. While paint types and chemistry can influence this durability, it is predominately the prevailing environment, the quality of application and the aspect of exposure the painted surface has to the environment, that dictates the degradation rate and ultimately the life of the coating.

Rarely do two different sides of a building or structure share the same prevailing environments. In buildings that are several stories in height, each elevation of approximately 10m presents a new set of environments. This pragmatic approach to structures and the increasing requirement for long term durability of 50 years or more, takes the process of management of an asset to a level of macro-management using risk based assessment concepts and techniques to determine the level of corrosion and maintenance. [3] [4] [5] [6]

The typical result is to assume that the worst conditions will prevail and conduct cost benefit analysis (CBA) to determine the value of patch painting on the worst areas until the rest of the asset degrades to a point where the economics of large-scale repair become viable. Given the rising cost of insurance, litigation and more restrictive and costly OH&S rules, the cost to conduct outside maintenance is escalating at a significantly greater rate than the consumer price index (CPI).

This paper will attempt to illustrate the economic effects and importance of treating galvanized steel as a composite material with varied options for design and maintenance depending on the prevailing conditions in use and the utilisation of particular assets.

The Hot Dip Galvanizing Process (AS/NZS 4680)

The application of a zinc coating onto steel members prior to installation at a site is normally achieved by dipping prepared steel into a bath of molten zinc. The resulting chemical reaction between the steel substrate creates a complex alloy of zinc and iron (steel), resulting in a thick protective layer which has distinct constituent layers.

The hot dip galvanizing process is relatively simple compared to most other corrosion protection systems and it is this simplicity that makes it economically efficient and technically effective. Most galvanizing that is specified for a long design life or for use in moderate to highly corrosive environments is performed to AS/NZS 4680 *Hot-Dip Galvanized (Zinc) Coatings on Fabricated Ferrous Articles*. [7]

The preparation process for steel prior to galvanizing involves the removal of scale, rust, oil paint and other surface contaminants. This is usually achieved through cleaning in a caustic solution followed by immersion in an acid bath. The steel is then rinsed in water prior to its immersion in a warm flux tank, usually made up of a 30% zinc ammonium chloride solution. The flux solution removes the oxide film that forms on the surface of the steel after the acid clean and also prevents further oxidation prior to the actual galvanizing process.

The galvanizing bath is made up of molten zinc and some trace alloys such as aluminium. The bath is normally heated in the range of 445°C to 465°C. On immersion in the galvanizing bath, the surface of the work is completely covered and the molten zinc reacts with the steel to form a series of zinc-iron alloy layers. This metallurgical reaction between the molten zinc and the steel is another factor in the reliability of galvanizing. The coating continues to form until the steel has reached the temperature of the molten zinc and the reaction is complete. The steel is then withdrawn from the bath at a controlled rate and an outer layer of molten zinc is carried with it. This solidifies to form the predominately pure outer zinc coating with its characteristic grey colour.

Because the metallurgical reaction and subsequent alloying of the galvanizing process can only take place once the steel has reached the temperature of the molten bath, the process can vary from several minutes for light articles to many times longer for heavy structural articles.

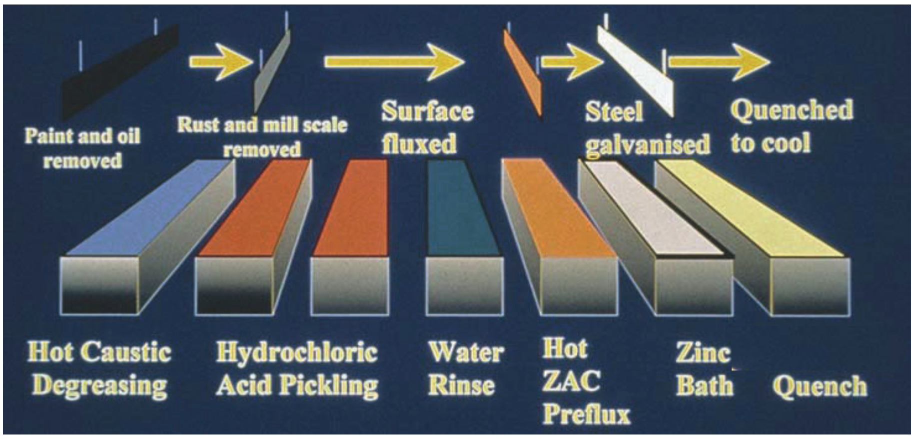


Figure 1: Hot Dip Galvanizing – General Process

2 PHYSICAL PROPERTIES OF GALVANIZED STEEL

The metallurgical alloying process of galvanizing provides galvanized steel with a number of unique properties. Since the process of galvanizing demands that the steel is fully immersed in the molten zinc bath, all surfaces of the work are completely covered, both externally and internally. This means that corners, edges, holes, welding seams and any recesses are completely covered and protected. This produces a uniform coating of zinc and zinc iron alloy layers over the complete article, no matter how complex in configuration, and specifically in areas that are potentially "corrosion hot spots" for other coating systems.

Secondly, because the zinc and zinc iron alloy layer thickness is determined primarily by the thickness of the steel, a standard minimum coating thickness is automatically applied to all areas regardless of orientation and access.

Figure 2: Large structural member: A large bridge beam measuring 12.5m in length, 1m through the web, and weighting 10.5 tonnes. This was a bridge beam for the State Rail Authority in NSW

Table 1 of AS/NZS 4680: Requirements for Thickness and Mass for Articles That Are Not Centrifuged

Steel Thickness (mm)	Local Coating Thickness Minimum (µm)	Average Coating Thickness Minimum (µm)	Average Coating Mass Minimum (g/m^2)
≤ 1.5	35	45	320
> 1.5 ≤ 3	45	55	390
> 3 ≤ 6	55	70	500
> 6	70	85	600

(AS/NZS 4680:1999 *Hot-dip Galvanized (zinc) Coatings on Fabricated Ferrous Articles* p8)

Corners and narrow edges actually produce a slightly thicker galvanized coating, thus providing greater protection from corrosion and impact on the areas most likely to suffer these. This contrasts with other coating systems, such as paint, which thin out in these critical areas and require expensive design and application modifications to improve their performance on such surfaces. Also, the predicted life to first maintenance of painted surfaces in AS/NZS 2312 applies to planar surfaces, not edges, crevices and other areas that may require further additional protective steps to improve their durability. [8]

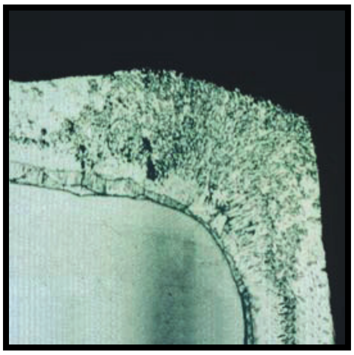


Figure 4: Micrograph of galvanizing on edge of steel

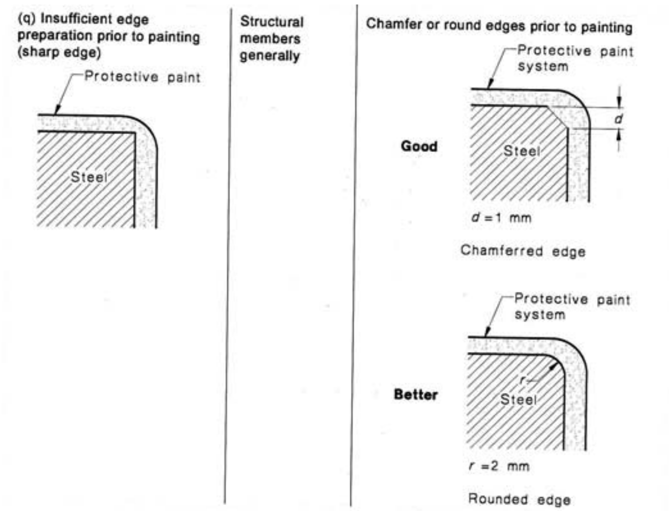


Figure 5: AS/NZS 2312 requirements for edge corrosion protection for paint systems. Note engineering design required to insure integrity and effectiveness of corrosion protection. This is not required for hot dip galvanized articles and results in less engineering expense.

The substrate of a galvanized section is nearly 100% iron while the most outer part or surface of the galvanized coating is 100% zinc. The intermediate layers have fixed ratios of zinc and iron progressing from ~25% Fe/ 75% zinc to ~10% Fe/ 90%

zinc, to 6% Fe/ 94% zinc, and finally pure zinc. These alloy layers can constituent more than 60% of the sectional thickness of the coating material.

The actually coating thickness is directly related to the substrate thickness, and thin substrates typically have thin coatings, while thick steels have thick or very thick coatings. Coatings are measured in either the thickness of the coating (µm) micro-metres, or by the weight per unit area g/m^2. The coatings found on thin material may be 100 g/m^2, whereas heavier sections of steel may be 600 g/m^2.

A feature of the galvanizing process is that the lower alloy layers are significantly harder than the base steel. In fact the delta layer can be around 50% harder than the base steel. This provides galvanized steel with much greater resistance to abrasion and mechanical damage. The advantage of this is not only seen in service, but also during transportation, handling and erection. It eliminates the need for expensive handling and transportation procedures and also the costly and labour intensive requirement of repair on site to insure the integrity of the corrosion protection system. These properties combine to make galvanized steel a legitimate engineering material that should be considered at the design and specification stage of a project. The correct design of a project can make the use of galvanized steel even more effective than just the usual expectations of its corrosion protection performance.

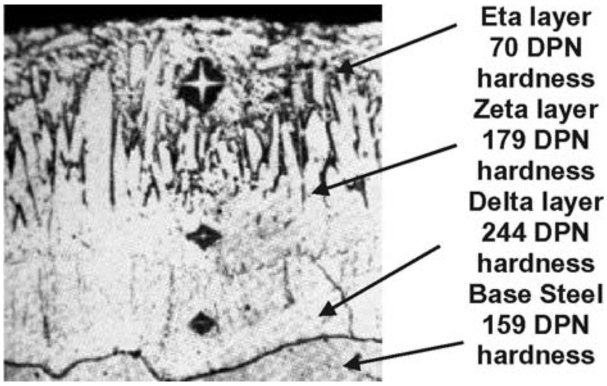


Figure 6: Galvanizing alloy hardness

3 HOT DIP GALVANIZING: PROTECTION MECHANISMS

Galvanized steel is protected from corrosion via a number of different mechanisms. As discussed above, it also provides physical protection in the form of greater abrasion resistance due to the increased hardness of the zinc iron alloys.

Barrier protection, as its name implies, works by providing an impermeable barrier over the steel item. Galvanizing provides barrier protection in two ways: firstly, the galvanized layer provides a protective physical envelope around the steel; secondly, the galvanized layer also develops a protective patina on its surface upon exposure to the environment. This is made up of insoluble zinc oxides, hydroxides, carbonates and basic zinc salts depending on the nature of the environment. Once the patina stabilises, it reduces the exposure of the base galvanized steel to the environment, thus considerably slowing the corrosion process. This patina regenerates itself after damage by very slowly consuming the zinc outer coating. Once the pure zinc of the outer layer has been consumed, the iron-zinc alloys are exposed to the environment and their corrosion resistance is up to 30% greater, providing even longer life.

The barrier protection qualities of galvanized steel are also enhanced by the fact that it is immune to ultraviolet radiation and thus will not degrade on exposure to Australia's harsh environment. Most other corrosion protection coatings will degrade on exposure to solar radiation. In fact, solar radiation is one of the key limiting factors to the performance of such coatings.

In the event of severe mechanical damage and exposure of the base steel to the environment, galvanizing also provides **cathodic protection**. The galvanizing performs in a similar way to other sacrificial protection systems, except in this case the sacrificial anode is distributed over the article to be protected and electrical continuity is assured. The cathodic protection characteristics of galvanizing ensure that mechanical damage does not result in the concealed under-film corrosion and potential catastrophic failure prevalent in some other protective coatings.

4 DURABILITY OF HOT DIP GALVANIZED COATINGS

There is a variety of data available from different sources on the durability of galvanized steel. Since the process has been utilised virtually unchanged since 1837, a large amount of empirical data has been recorded from case studies in a large number of operating situations, locations and corrosive environments. In fact, one of the major advantages of hot dip galvanizing to design and specification engineers is its extensive history of use and the fact that much of this is on record. This history of use also assists maintenance engineers in planning for infrastructure life, maintenance and remediation. The corrosion rates for galvanized steel in a variety of situations are readily available and these can be used to efficiently and accurately plan maintenance and asset life cycles.

Engineers can choose to use empirical data from selected sources such as Porter (1991), Slunder & Boyd (1983), Porter (1994) and Zhang (1996) or refer to standards that address corrosion rates for various metals and corrosion protection systems such as ISO 9223 and AS/NZS 2312. [9] [10] [11] [12] [13]

Table 2 Corrosion Rate of Galvanizing on Vertical Surfaces in Various Environments (Note 1)

ISO 9223 Corrosion Category	AS/NZS 2312 Corrosivity Description	Typical Environment	Average Zinc Corrosion Rate µM/year
C1	Very Low	Few alpine areas, dry interiors	< 0.1
C2	Low	Arid/rural/urban, Interiors - occasional condensation	0.1 - 0.7
C3	Medium	Coastal High humidity interiors	0.7 - 2
C4	High	Sea-shore (calm) Swimming Pools	2- 4
C5	Very High	Sea-shore (surf)/ offshore	4 - 8

Note 1: It should be noted that the corrosion rates given in this table provide only generalised guidance and other factors, such as specific microclimates and unwashed areas, may have an influence.

AS/NZS 2312 provides advice on the use of various corrosion protection systems including galvanizing. This standard is also useful to the design engineer as it highlights design issues to be considered when using different systems.

5 DUPLEX SYSTEMS FOR HOT DIP GALVANIZING

In circumstances where the galvanized coating may be attacked by a particularly corrosive environment, the application of a suitable inert coating which insulates the zinc from that environment will prolong the life of the galvanizing. Usually this coating is either paint or powder coating. The main areas of significance are in highly coastal or marine service, in areas of high and persistent rainfall and in industrial locales. It is worth noting that the impact of industry in Australasia, in particular acidic fallout, is low and has diminished considerably over the last 40 years with the demise of high sulphur fuels and the development of environmental protection agencies. [14] Notwithstanding, there remain many situations where painting enhances the service life of galvanizing and the synergistic increase in service life of duplex coating systems in high corrosivity locales has been well documented in such situations. [9]

The requirement for enhanced durability not only applies to a general locale or macroclimate, but also may apply to a quite specific microclimate. In particular, areas exposed to airborne contaminants, but protected from the cleansing influence of rainwater, can be considerably more vulnerable to corrosion than areas on the same structure exposed to the elements.

The corrosion rate of well-applied duplex coatings on washed surfaces in use is normally greater than the sum of their separate life expectancy. Typically if applied in severely polluted climates, the increase factor for a duplex coating is 1.8 to 2.0 times the sum of the life of each system, in sea water 1.5 to 1.6 and in rural climates 2.0 to 2.7 times. [9]

Another advantage of painting over galvanizing is that the time to first maintenance can be deferred well beyond the criterion for first maintenance of steel over-coated with organic paints alone. Specifically, AS/NZS 2312 sets the criterion for the maintenance painting of steel when "about 0.2- 0.5% of the steel in any particular area is showing signs of rusting". For painted galvanizing, the threshold for maintenance painting is set at 2%. This is because an effective duplex system resists corrosion beneath the paint finish. Without the anodic influence of zinc, corrosion of ungalvanized steel can proceed under the paint film before it becomes visible on the surface. In addition, when breakdown of the metal substrate occurs, the aesthetic impact of the whitish zinc corrosion products tends to be less than the red/brown iron stains of non-duplex systems.

The main criteria for painting galvanizing, such as preparation techniques, paint selection and paint application are the same as for painting bare steel.

Careful selection of paint systems is crucial in high corrosivity environments. An effective corrosion inhibitive primer and adequate total paint thickness is usually a prerequisite. Indeed, many failures of paint systems over galvanizing in arduous service conditions can be traced to either an inappropriate primer or insufficient paint thickness. [15]

Examples of the durability of such duplex systems abound.

Figure 7: Duplex system used in high corrosivity conditions in Spencer Gulf

Figure 8: 130 year old galvanized telegraph pole recovered in Far North Queensland

5.1 Hot Dip Galvanizing and Infrastructure

The oldest known galvanized structures still in existence in Australia are telegraph poles that were used in the first Morse Code line in far northern Queensland. Towers recovered from the Charters Towers line are over 130 years old and have survived in the moist tropical environment with their galvanized coatings still above the thicknesses demanded by the current Australian Standard. The coatings are expected to last into the next century!

The use of galvanizing in the electrical industry is extensive and has a long history. In many areas, galvanized tower construction is considered to outlast the economic life of the system and such towers are more likely to be replaced due to system redesign rather than degradation. In many power companies, the proven performance history of hot dip galvanizing has led to minimum design life requirements to first maintenance of at least 50 years.

There are transmission towers in Victoria in environments of varying corrosivity that are still in operation approximately 80 years after commissioning. Surveys of electricity transmission companies indicate that towers in even the most severe coastal exposure are achieving 30 years of life without any maintenance or before any sort of maintenance is even investigated or considered. Many Australian companies now base their system designs on an average life of 50 years for galvanized structures.

Because the galvanizing process has been around for many years, this provides a significant resource in understanding the relationship or correlation between environmental conditions and corrosion rates of galvanized steel [16] [17] [18] [19]. A fortunate relationship prevails for most industrial, urban, coastal and rural environments, in that the corrosion rates of the zinc is significantly different at each site but after the first year of exposure, is typically a constant corrosion rate. That is, as constant as the average environmental factors such as meteorological conditions over the year can be.

In laboratory studies, similar consistencies in the corrosion rates were found leading to the widely held concept that corrosion rates are steady. In fact this is actually conservative, as it has recently been shown that the development of the complex corrosion products on the zinc surface physically and chemically restrict the chemical process of corrosion and slow the rate significantly. [20] [21] [22]. So the corrosion rate of older steels with well established corrosion products tends to be lower. It is interesting to note that the development of corrosion products as time goes by, in the vast majority of environmental conditions, actually improved the performance or corrosion resistance of the galvanized coating system. However, this “linearity” of the sacrificial consumption of galvanized coatings makes maintenance planning a much simpler task and reduces the complexity and onerousness of the inspection process. An added feature of such systems is that even after most of the galvanized coating has sacrificially been consumed, the lack of under-film corrosion means that the exposed steelwork is basically in its original condition.

The remote location of many electrical installations and associated infrastructure also places great demands on the transportation of steelwork that will minimise damage to corrosion protection coatings. The robustness and increased abrasion resistance of galvanized steel makes it particularly suited to such applications.

5.2 Other Coating Systems

5.2.1 Electroplated zinc

A common coating system is electroplated zinc onto the steel as found on light weight components, often suitable only for indoor use due to the thin zinc coating and resulting readiness to corrode after a relatively short exposure outdoor. The coating method does not promote the development of a complex layer system, more than a few atoms in thickness, between the coating and the substrate, and can be considered as an added layer.

5.2.2 Paint Coatings

Paint coatings offer protections basically by providing a barrier protecting the substrate from the environment, or as a barrier with anodic or corrosion protective flakes of metals in the coating. The coating does not combine with the substrate but is conformal and takes the shape of the substrate.

Paint systems do have ranges of expected performance or durability for different environmental conditions, of which the most significant are thermal, moisture, and UV[23]. The paint application process requires significant surface preparation [24] [25] which needs to be done correctly if adhesion to the substrate is to be maintained over the life of the coating. Painting is a logistical and time consuming process for protection of 5-20 years (environment dependant). This significant contributor to the cost for maintaining structures is also recurrent and painting effectively locks the maintenance program into a cycle where patch painting or full painting, which are more economically viable than excessive surface preparation. This assumes the management of the asset is proactive in maintaining the condition of the asset.

6 GALVANIZING AND ASSET MANAGEMENT

The galvanized zinc coating is significantly different in performance and the way it is bound to the substrate, compared to electroplated or painted coating systems. This distinct difference can be best described as a type of layered or laminar composite system of several Zn/Fe alloys.

The broad performance expectations of the coating in numerous general environmental conditions is known and subsequently the performance can be readily predicted, as noted in AS2312 Standard [8] [26].

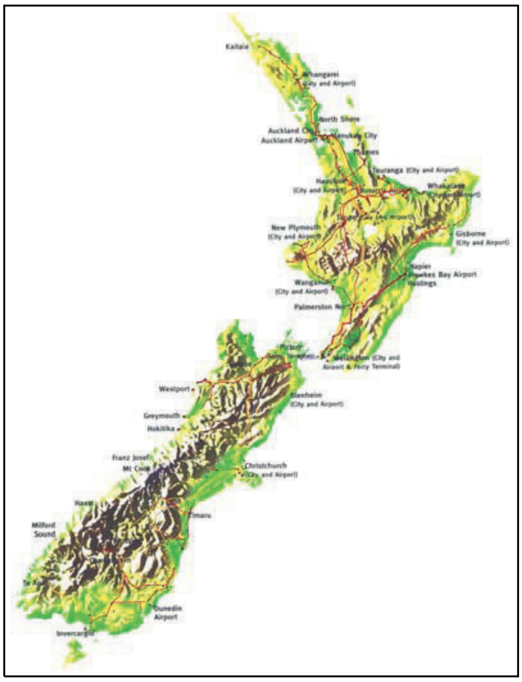


Figure 1: Power Distribution lines in New Zealand [27]

In the following section, the economics and issues of management are discussed with reference to the power distribution towers in New Zealand. This group of assets number more than 25,000 and are spread throughout the country with particular higher intensity in industrial and urban regions of Auckland. In addition the major power source for New Zealand is the hydro power generation plants in the alps of the South Island, while the majority of the population live in the warmer climes of the top of the North Island. This fact has resulted in very long tracts of transmission towers running in a South to North direction.

6.1 Lattice Tower for power distribution in New Zealand

The use of galvanized steel in lattice tower construction, in comparison to other coatings has several advantage and empowerment of best practice management procedures which are unique to this coating.

The design criteria of the asset is to have a life of 50 years and to be in a condition at the end of that period which will permit refurbishment. Basically, the structural steel would need to be in present at the end of the period such that a paint or other coating can be applied.

The maintenance management have several options;

1. No maintenance - Coat with zinc and do no maintenance.
2. Reactive maintenance – Coat with zinc, occasionally monitor and repair when they are in a serious structural condition that may impair function
3. Proactive maintenance – Monitor regularly and maintain with spot repair if economically viable or proceed with full maintenance.

On close inspection it becomes apparent that to achieve the durability objective, one or more of the three management styles can be applied to each of the environmental conditions. If the primary goal is to have the asset, at the end of 50 years, to be as structurally sound as it was at the time of construction, then it is possible to identify the benefits and limitations of the management styles.

Figure 2: Heavy duty High Voltage Tower in Christchurch (NZ).

Figure 3: Corroded tower with some rust of steel (dark red) in USA.

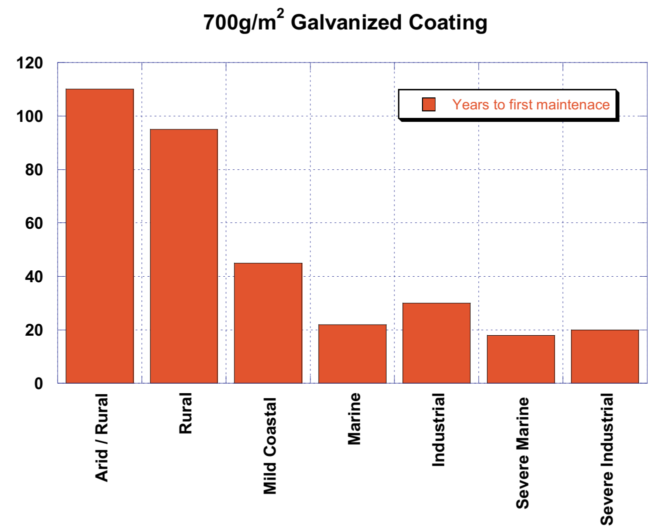


Figure 4: Corrosion rates of zinc (100gm/m^2) in various Australian environments [28]

6.2 No Management of Towers

With reference to Figure 4, operating a No Management style, which is surprisingly common, would work for all lattice towers in the environmental conditions of arid/rural, rural, and possible some location with very mild coastal exposure. Towers located in all other environments would very likely be severely corroded and exhibit a high risk of failure. The economical advantages of this style are in the significant cost reduction of the need for maintenance but the down side is that there is likely to be a significant shortfall in skilled persons when periods of high demands are required. The economic rationalists can argue that the industry has paid for the protection of the tower when the items were galvanized and the best use of that resource is to let it all corrode away. However the maintenance contractors readily note that the highest cost in painting is the surface preparation. In fact it is currently 8 to 10 times more costly to prepare rust with a grit blast than to prepare aged zinc with a high pressure water blast.

6.3 Reactive Management

The Reactive Management style or commonly referred to as 'fighting spot fires', can maintain the current conditions with arid, Arid/Rural, Mild Coastal, and Industrial locations, but is likely to be swamped by demand from the Marine and Severe environments. In economic terms, there is a fine balance between programs that are heavily subjected to spot repair processes such that it becomes cheaper to do a total repair on a lattice tower.

6.4 Proactive Management

The Proactive Management style is ideally suited for the management of lattice towers which are located in the high corrosion areas and by default, the inclusion of the towers in other environments would be relatively easy. The management of individual towers with respect to coating performance is based of the premise that the towers are inspected and the conditions are reported at least once every 10 years. The predictability of the degradation of galvanized coatings allows forward planning and management of the limited financial resources to achieve the objectives.

6.5 Maintenance of Galvanized Structures

One of the issues in maintenance is the cost effectiveness of maintenance. If we were without an policy of zero structural loss after 50 years, then such a policy would dictate the level of management of the assets. In the power industry, the periods of outages in the supply of electricity potentially are a very costly business and the government regulators are very keen to keep these to a minimum. Without considering these objectives, there are four procedures that can be implemented at each tower. These are:

1. Complete Replacement:

This complements the 'No Maintenance' program and this has significant merit if in isolated or difficult to access locations, and if only a short life time is required. The lattice tower is made from thin sections (braces) and heavy sections (legs) and given the elevation and different orientations, no two sides or elevations will have the same corrosion rates of the zinc. It is common that significant replacement of light section steel, bolts and some patch painting would occur prior to the cost effectiveness of total replacement becomes apparent. In favour of total replacement which has the benefits from having new zinc (galvanized steel) and subsequently many years before first maintenance.

2. Painting after galvanizing fails:

Paint galvanized steel after the occurrence of the deep red steel corrosion product indicates that some loss of steel section has occurred and that the surface preparation re3quires the use of the grit blasting process. The relatively high cost of this process and the physical issues for the contractors, including a significantly high OH&S risk level, puts this as a marginal option. The estimates [29] put the rust removal process increases costs by at least 20% (2004/5). The CBA for this method may be very close to the price for complete replacement of a tower, and the benefits from a new tower is the new thick zinc coating (galvanizing) and many years before first maintenance.

3. Painting before the galvanizing fails

The rationale of painting galvanized steel before the appearance of sheets of red rust is to obtain the full benefits of the zinc coating and to minimize the costs in surface preparation. This procedure is also combined with a replacement of light steel section and fasteners (where needed). This methodology fits well with the proactive management process as there are effectively only small windows of 2-3 years for painting in moderate to severe environments.

4. Duplex System – Paint over new galvanized steel

The application of a high performing coating system over very quality galvanized steel has economical advantages where the lattice tower is in a very corrosive marine or industrial location., where typically the tower life would be relatively short. The cost of painting is at the lowest and the quality is at the highest since the operations are typically inside and fully access to all surfaces is possible. The CBA would be favourable again in locations where access is difficult and a history of high corrosion.

Costings

A recent publication (Nov 2005) [29] provided details on the costs of tower painting and tower replacement for a typical double circuit 220kV tower with a surface area of 420m^2.

The following details may provide an insight into painting a tower with easy road access and in a moderately corrosive environment [29].

Painting		CBA	
(220kV Double circuit tower)		Initial asset value	NZ$60,000
Surface preparation and painting	NZ$49/m^2	Discount rate	7.5%
Rust removal	NZ$70/m^2	Period	100 years
add zinc rich primer	NZ$25/m^2	Annual patrol cost	NZ$100
		5 yearly condition assessment	NZ $250
		10 year routine maintenance	NZ $2,000
		Painting maintenance after 15 years	NZ $8,000
		Similar condition at end	

Table 1: Costing Values for lattice towers in New Zealand (2004/5) [29]

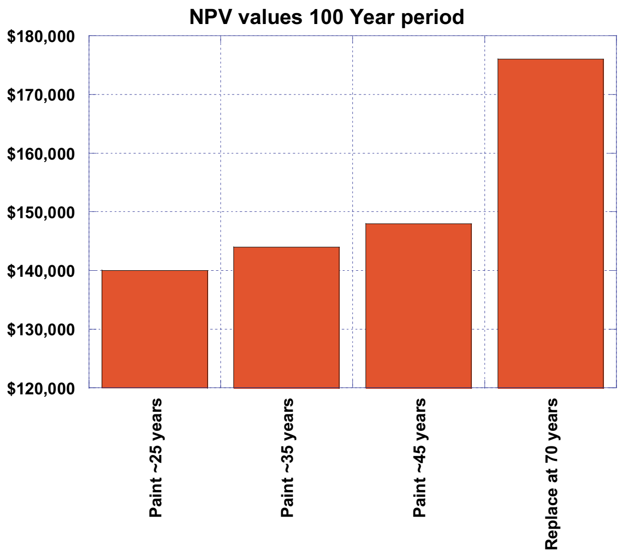


Figure 5: The NPV for the options of painting or tower replacement [29]

7 CONCLUSION

The hot dip galvanizing process has performed effectively in a wide range of conditions and applications for over 150 years. As the cost of steel, infrastructure and labour increases, it is essential that engineers consider their economic responsibilities both to the public and the private institutions that use, operate and invest in such infrastructure. Hot dip galvanized steel provides an ideal solution as an engineering material since not only does it protect the increasingly valuable underlying material of steel, but it also reduces the need for maintenance, costly inspections and potential catastrophic failure due to under-film corrosion.

The proper use of galvanized steel requires some understanding of design requirements and total cost analysis of its advantages over the life of a structure. Life cycle analysis and the costs associated with the choice of design and engineering materials are becoming increasingly relevant to the way that infrastructure is planned and designed. The galvanizing industry is currently working on developing further rigorous life cycle analyses for use by industry. These will include the utilisation of existing case histories to make the information as relevant and practical as possible.

The alloys layers that exist on galvanized steel set this coating apart from all other coatings which are adherent or at best form weak chemical bonds. This fact makes the concept of galvanizing as a layered or lamella composite of zinc rich alloys, is an apt descriptor.

Galvanizing offers substantial corrosion protection performance that can well exceed 50 or even 100 years and is related predominately to the level of exposure to the prevailing environmental conditions (temperature and moisture) and includes pollutants such as marine salts and industrial types.

It has been reported that it is prudent to extract the maximum life out of the zinc coating on a lattice tower and before there is any requirement for abrasive blasting, the tower should be painted. This is significantly cheaper over 70 years compared to replacing the whole tower.

8 REFERENCE LIST

1. Biezma, M.V & San Cristobal, J.R. Methodology to Study Cost of Corrosion. Corrosion Engineering, Science & Technology, Volume 40, No. 4, Dec 2005 pp 344-352
2. Koch, G.H., Brongers, M.P.H et al Corrosion Costs & Preventive Strategies in the United States. CC Technologies Laboratories for Office of Infrastructure Research & Development, Report FHWA-RD-01-156, Sept 2001
3. Moss C.J., Scully M.J., Camp A., and Joyce M. Risk directed inspection. Corrosion and Materials. 39, Paper 051. 1999.
4. AS/NZS 4360: 1999 Risk assessment
5. Braithwaite s., Faulkner J.R.E, and Calles R.E. Lattitude and altitude: Safety in Australia. National conference of the Risk Engineering Society. 1997.
6. Dinon M., Atkinson B., and Patch B. Risk management and the process industry corrosion engineer. *Corrosion and materials* 26: S3-5, 2001.
7. AS/NZS 4680:1999 *Hot-dip galvanized (zinc) coatings on fabricated ferrous articles*
8. AS/NZS 2312:2002 *Guide to the protection of structural steel against atmospheric corrosion by the use of protective coatings*
9. Porter, F *Zinc Handbook: Properties, Processing and Use in Design*, Marcel Dekker, NY, 1991
10. Slunder, C & Boyd, W *Zinc and its Corrosion Resistance*, ILZRO, NY, 2nd Ed. 1983
11. Porter, F C*orrosion Resistance of Zinc and Zinc Alloys*, Marcel Dekker, NY, 1994
12. Zhang, X *Corrosion and Electrochemistry of Zinc*, Plenum Press, NY, 1996
13. ISO 9223*Corrosion of metals and alloys – Corrosivity of atmospheres – Classifications*
14. Bartlett, D *Industrial Pollution and Its Impact on Corrosion and Corrosion Mitigation Practices*, Australian Corrosion Association Conference, Nov. 2001
15. Pimentel, E, Bartlett, D & Thomson, G *Merits & Myths: Galvanized & Painted Protective Coatings for Steel*, 2005
16. Vazquez, A. J and de Damborenea, J. J Corrosion resistance of galvanized coatings: retrospective evaluation. 16th International Galvanizing Conference. GB4-1-GB4-12. 1991. Spain.
17. Spence, J. W., F. H. Haynie, F. W. Lipfert, S. D. Cramer, and L. G. McDonald. Atmospheric Corrosion Model for Galvanized Steel Structures. *Corrosion* 48: 1009-1019, 1992.
18. Legault, R. A. Atmospheric corrosion of galvanized steel. Ailor, W. H. Atmospheric Corrosion. 607-613. 1982. New York, John Wiley & Sons.
19. Sheehan, R. Galvanized coating life predictions for industrial projects. Corrosion Management 8(1), 21-22. 1999. Australia.
20. Odnevall, I. and Leygraf C. Formation of $NaZn_4Cl(OH)_6SO_4 . 6H_2O$ in a Marine Atmosphere. Corrosion Science 34(8), 1213-1229. 1993. Great Britain, 1993 Pergamon Press Ltd.
21. Odnevall, I. and Leygraf C. The formation of $Zn_4Cl_2(OH)_4SO_4 \quad 5H_2O$ in an urban and an industrial atmosphere. Corrosion Science 36(9), 1551-1567. 1994. Great Britain, Pergamon / Elsevier ScienceLtd.
22. Leygraf. C and Graedel T.E.The atmospheric corrosion of zinc. Atmospheric corrosion. New York, Wiley-Interscience, John Wiley & Sons. 2000, 329-340.
23. Hare, C. H. Irreversible Degradation of Paint Films. *Trouble with Paint* 26-34, 1999.
24. Bullett, T. R. and Prosser J. L. Factors affecting the adhesion of paint films to metal substrates. 8th FATIPEC Congress . 374-381.
25. Mattox, D. M. Adhesion and Surface Preparation. Bunshah, R. F. DepositionTechnologies for Films and Coatings Developments and Applications . 63-83. New Jersey, USA, Noyes Publications.
26. Francis, R. A. AS/NZS 2312: A New Standard in Coating Selection. *Corrosion and Materials* 20: 5-9, 1994.
27. Jacobs, T. E. and Haberecht, P. W. New Zealand: Environmental Diversity and Power Lines. CIGRE. 1-11. 1999. Australia.
28. Galvanizers Association of Australia
29. Boardman M. and Marshall W. Galvanized coating on power equipment: Optimizing life cycle costs. CAP 05. Paper 34, 8pgs. 2005.

ASSET MANAGEMENT OF MEDIUM VOLTAGE CABLE NETWORKS

Birtwhistle D [a] , Gilbert R [b], Lyall J [a], ‘ Oyegoke B [a], Powell L[b] and Saha T [c]

[a] *School of Engineering Systems, Queensland University of Technology, George Street, Brisbane, Queensland 4001,*

[b] *Ergon Energy, Queensland, Australia. c University of Queensland, Brisbane, Queensland, Australia.*

Abstract: This paper describes the background to research currently underway associated with asset management (AM) of degraded cable networks in North Queensland. In the paper the main failure mode of cables installed twenty years ago is identified as the growth of water trees through the cable insulation. The costs of managing water tree degradation in a large distribution network are considered and the economics of cable replacement in a regulated monopoly are examined. Condition monitoring techniques for assessing the condition of cables are reviewed and the possibility of silicone liquid injection as an alternative to cable replacement is discussed: a research methodology to support improved cable AM is also presented.

Key Words: XLPE Cables, Water Trees, Cable Refurbishment, Insulation Diagnostic Testing

1 INTRODUCTION

High voltage underground cable systems utilizing XLPE insulation were developed in the United States in the 1950s. After about one decade of service insulation failures started to occur that were found to be caused by the ingress of moisture into the cable insulation. Considerable research into this phenomenon has been conducted over many years and it has been established that moisture is induced into the cable insulation by the a.c. electric fields. Water progresses across the insulation under the effect of the electric field in microscopic channels with frequent branches. These water channels are referred to as “water trees”. The main factors contributing to the initiation and growth of water trees have been established and a range of diagnostic techniques for the identification of the presence of water trees in cables have been examined. In recent years new types of cables have been developed that effectively preclude moisture or resist the encroachment of water trees, and techniques have been developed by a few companies for removing moisture from cables in-situ and filling the vacant channels in the insulation with insulating fluids, thereby restoring the cable insulation.

Water trees had not been discovered in Australian cables until repeated failures of 22 kV cables in North Queensland were attributed to the presence of water trees. Subsequent investigations suggested that the conditions leading to water tree development occurred in a significant proportion of existing cables. There was therefore a requirement to establish procedures for managing degrading cable networks to maintain high reliability whilst at the same time keeping down costs. To do this it is necessary to effectively utilise the results of the very considerable body of research that had been done on water trees over several decades. In this paper we consider fundamental factors influencing asset management of degrading cable networks and describe research currently underway that is directed towards to proving information needed to improve decision making in allocation of resources.

2 FAILURE MODES OF DEGRADING CABLES

2.1 Water Trees in XLPE Cables

In the cables investigated there is one cable per phase in three-phase systems with a voltage between the core of the cable and sheath of $22/\sqrt{3}$ kV. Cables are either buried directly in the ground or are pulled through buried ducts. Moisture either permeates through the insulating sheath or is drawn into the intestacies of the stranded core. Water trees in the insulation may be observed by making thin transverse slices of the recovered insulation and staining the thin sample: water trees observed microscopically in a sample of insulation taken from [1] are shown in Figure 1.

As water trees grow through the insulation electrical trees with carbonised tracks may develop at the tips of the water trees. Electrical trees may cause very small transient currents through the cable insulation that are known as partial discharges (pds). Such discharges are not expected with water trees alone.

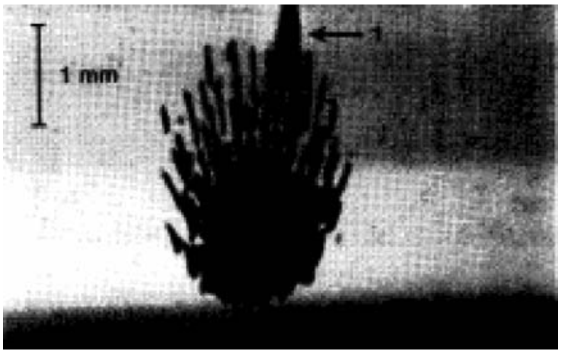


Figure 1. Vented Water Tree in a Polymeric Cable [1]

When trees completely bridge the cable insulation a high current may flow and an electric arc may be established causing damage to the cable before the electrical supply to the cable is automatically disconnected. The damaged part of the cable needs then to be repaired. First the exact position of the fault has to be identified by one of several methods. For direct-buried cables the faulty part of the cable must be excavated, the damaged section of the cable must be replaced and a new length of cable inserted using two high voltage joints. In some cases location of faults is difficult using existing technologies and it may be necessary to purposely cause further damage at the fault point to establish a good connection to ground to assist fault position location: this is done by connection of a high-voltage/high-current supply to the damaged cable to burn out the fault – this process is known as "thumping" the cable. Cable repair may take several days during which period customers fed by the cable may be unavoidably deprived of electricity supplies.

2.2 Other Cable Failure Modes

The insulation quality of a cable system is affected not only by the insulation of the cable itself but also the insulation of joints and terminations. At the time of installation of the original cable, lengths of cable are joined by special high voltage joints that are buried in the ground: over time additional joints may be introduced when cables are repaired as described above. Special insulating terminations are required at the ends of the cable where the high voltage conductors are connected to other equipment. Faulty materials or poor workmanship may lead to gradual degradation of the insulation of both these types of cable ancillaries.

Failure of the insulation may also be accelerated by abnormal voltages and currents occurring in the cable network. For example voltage surges due to lightning or switching, whilst limited by protective equipment such as surge arresters may trigger the failure of degraded insulation. Overloading of the cable for short periods will increase the cable operating temperature and may also accelerate insulation failure.

3 ECONOMIC CONSIDERATIONS

3.1 Costs and Regulated Income

Cables of the type discussed above are part of electricity distribution networks. Currently companies operating distribution networks in Australia may be privately- or government-owned monopolies the income of which is fixed by rules formulated by independent government-appointed regulators. In Queensland rules are fixed for distribution network companies by state regulations [2]. In order to examine the factors affecting decisions regarding the management of cables it is necessary to consider both the costs and the income from this class of deteriorating assets.

The costs of operating a cable with degrading insulation from the present time can be considered as being:

(a) Costs of operating the cable until replacement or refurbishment

Costs are mainly those associated with making diagnostic tests to ascertain the condition of the cable and costs of repairing intermittent failures of the cable insulation. The costs of cable repairs at this time comprise only the costs associated with locating faults and jointing in a length of new cable. In the regulatory regimes currently in force in most Australian states for electricity distribution there is no effective penalty for prolonged loss of supply to customers due to network faults.

(b) Costs of replacement/refurbishment plus operating costs of the new/refurbished cable.

Costs of cable replacement include: excavation, removal and disposal of the deteriorating cable; purchasing of the new cable; laying, jointing and reinstatement costs for the new cable. Refurbishment of the cable by silicone fluid injection will include cots of the injection and replacement of terminations and joints by special equivalents that allow flow through of the silicone fluid. Once cables are replaced or refurbished there are on-going operational costs of the type described in (a) above.

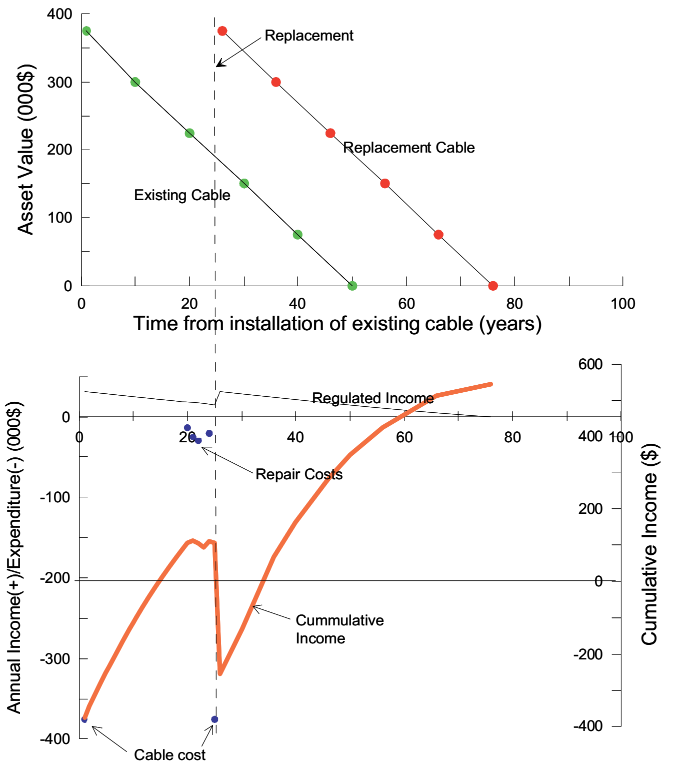


Figure 2. Costs and Income of a 1 km 22 kV Cable

Regulated income to the cable owner is based on a formula that includes components for capital and operating expenditures. The capital-related component is estimated from a depreciated optimised replacement value (DORC) of the network multiplied by a weighted average cost of capital (WACC). Individual groups of components of networks such as cables are ascribed an average lifetime and depreciated by straight line relationship. Cables have been a long-life asset and lifetimes of 40 or 50 years are generally fixed for all cables. Each cable circuit can therefore be regarded as producing income as well as having operating costs. In Figure 2 we have illustrated this situation for a hypothetical 22kV cable circuit of length 1 km. The upper graph in Figure 1 shows the depreciated values of the original cable installed in year zero and a replacement cable installed in year 26. The lower graph in Figure 2 shows annual income and expenditure associated with the cable circuit with all amounts being expressed in today's dollars. In Figure 2 large negative expenditures are shown in year 0 and 26 for cable purchases. Accounting practice also requires that the "book value" of the replaced cable should also be deducted at the time of replacement. This is not shown in the Figure which includes only "real costs'. Typical costs of repairs are shown between years 20 and 26 and income generated with a WACC of 8.5% is also included. The total cumulative income from the cable circuit is also shown in Figure 2 over a 76 year period. It can be seen that after about year 14 the cable circuit begins to generate a financial return but failures followed by replacement of the cable delays return on capital invested until about year 33. This is provided that there are no further failures of the new cable.

In Australia DORC of distribution networks are periodically adjusted upwards by the regulator about every five years [3]. This adjustment has not been at the same rate as changes in the CPI. It is therefore desirable when making long-term comparisons to estimate the net present value of changes in the future income from the circuit when making decisions about replacement.

3.2 Community Costs

Whilst at this time there are no financial drivers for cable owners to take into account prolonged loss of electricity supply that may be experienced by customers supplied by circuits during repair periods, there are clear community pressures to try to minimise electricity "outages". Research into costs of unavailabity of electricity indicates that, whilst the current cost of electricity to the domestic customer is about 0.1$ per kWh, the cost of the unavailability of power to the community may be of the order of $10 per kWh. On this basis the loss of an average load of 1 MW for a period of only two days would incurr a community cost of almost $500,000: this dwarfs the repair costs shown in Figure 2 and emphasises the undesirability of permitting repeated cable failures.

3.3 Network Expenditures

In ageing cable networks there may be considerable numbers of cables at risk of insulation failure possibly due to deterioration of cable insulation, water trees or to poor installation practices. In network companies there will be only a limited amount of capital available annually for cable replacement and consequently there is a need to prioritise replacement of degraded circuits. This needs to be done in such a way as to minimise disruption to electricity supplies whilst at the same time maximising the overall cumulative income for the cable network. In the next section we discuss current research activities that are directed towards realising these objectives.

4 ASSET MANAGEMENT RESEARCH

4.1 Techniques for Identification of Degraded Cables

Cable failures in the Ergon Energy 22 kV network have been attributed to both water trees in cable insulation and failure of the insulation of underground joints. The identification of the cause of failures requires forensic evidence from the point of failure and other part of the failed cable. Generally the site of the failure suffers damage at the time of failure but using well-established procedures [4] adjacent parts of the insulation may be microtomed, stained and microscopically examined to reveal pictorial evidence of water trees of the type shown in Figure 1: the condition of the cable may be established by the tedious process of sampling insulation to estimate the number and length of water trees present in the cable insulation.

The identification of the state of cable insulation from non-invasive electrical measurements has been the subject of research for many years and several useful methods have been developed. High voltage "Withstand Testing" with Very Low Frequency (VLF) (typically 0.1 Hz) requires large and expensive equipment. This type of testing produces similar dielectric stresses to power frequency tests at the same voltage and a VLF voltage of 2.5-3.0 rated voltage is applied for up to 60 minutes [5]. Several companies manufacture VLF units. Only a single result is obtained from this test and it is not possible to obtain information on the degree of ageing and continued serviceability of the cable. Perhaps the most interesting diagnostic techniques for overall condition rely on changes occurring in the dielectric properties of the cable insulation as a result of in-service ageing due, for example, to water treeing and useful results have been obtained from the dielectric response of cables measured in either the time domain or frequency domain. Measurement of time-domain polarisation/depolarisation current (PDC) or return voltage following application of a dc voltage have been used. Hoff and Kranz [6] have developed a system using Isothermal Relaxation Current (IRC) analysis tools for the determination of the ageing status of polymeric cables and they claim that their system can assess residual lifetime and residual insulation strength in a quantifiable way. IRC system is now commercially available.

Frequency-domain spectroscopy (FDS) [7] facilitates measurement of permittivity and dielectric loss factor in the frequency range 0.1 mHz to 1kHz. A sinusoidal voltage is applied to the cable and the complex dielectric constant is determined from the amplitude and phase of the resulting current. Variation of these parameters with voltage and frequency provides insights into the condition of cable insulation and it has been shown [8,9] that the features of the dielectric response obtained from FDS correlate well with water-tree content and breakdown voltage. When the dielectric response increases more than proportionally with increasing test voltage it has been suggested [10] this may flag the presence of long water trees. As the magnitude of the breakdown voltage in degraded XLPE cables is related more to the length of the water trees than their number[10] it is possible that FDS may prove to be a useful diagnostic method. Werelius et al [7] have introduced a classification of different types of response and relate these to the degree of water-tree deterioration of the cable insulation: this research on dielectric spectroscopy has led recently to the development and manufacture of a commercial instrument incorporating a HV supply. Hvidsten et al found that the condition of old cables with a high density of vented water trees was correctly assessed by all test methods [11]. Their results indicate that in order to avoid dielectric breakdown during diagnostic study, the test voltage should be below twice rated voltage. They also noted that the response from installed accessories should be known in order to avoid wrong interpretation of cable condition. Hoff and Kranz [12] however were unable to draw any definite conclusions from their work as they felt that the influences of other factors due to the environment and pollution needed explanation. At this stage dielectric response methods provide interesting possibilities for detecting the condition of water treed cables but further work needs to be done to build up confidence in the use of these techniques for lifetime prediction of degraded cables. Methods of monitoring cable condition based on the measurement of partial discharges are well established. At rated voltage even cables with long water trees may not produce pds. Techniques have been developed that involve the application of higher test voltages for a short time (0.3 – 1.0 sec) with measurement of the associated pds [13]. There are, however, concerns that even the temporary overvoltage may cause failure of the cable. Several workers including colleagues at University of NSW [14] have developed "on-line" systems for the measurement of pds and that might provide early warning of cable failure but at this stage correlations between measurements from such systems and cable failures need to be established.

4.2 Cable Replacement and Refurbishment Options

A silicone fluid known as CableCure™ manufactured by Dow Corning has been developed for treatment of XLPE cables with water trees. The fluid is injected into the stranded conductor of cables. It slowly impregnates the insulation and bonds with water molecules, drying the cable, and filling microvoids. It is claimed [15] to heal water-tree damage and retards the growth of new trees. The breakdown strength of the cables improves by around one-half percent each day, continuing to increase for up to two years as impregnation continues. The cost of refurbishment was found to be 30 to 50% of the cost of replacing cables

Electrical utilities in Oregan [15], New York [16] and Norway [17, 18] all report very good experience with medium voltage cables injected with CableCure™ . The treatment is shown to give significant increases in 50Hz and impulse breakdown voltage of the water treed cables and one utility reports reduction of failures from hundreds to fifty. A 1994

UTILX/Dow Corning paper [19] reviewed the use of CableCure by utilities in North America and Europe up to that time. Acording to this paper "silicone dielectric enhancement technology" was a reliable alternative to cable replacement for failing underground cable rated at 15-25 kV. Problems were ascribed to under-penetration and non-uniform distribution of fluid through the insulator but new silicone fluids were claimed to have largely overcome these problems. Independent researchers [17] concluded that: CableCure restoration strongly increases the residual 1.2/50 μs lightning impulse breakdown strength of water-tree aged XLPE cable but CableCure must remain in contact for a long time in order to obtain a permanent effect. On the basis of previous experience with CableCure™ silicone dielectric enhancement technology for cable refurbishment appears to be an attractive alternative to cable replacement. It is known however that the progress of water trees is profoundly affected by cable temperature and from the foregoing it can be seen that all documented experience to date has been in somewhat colder climates than North Queensland. It is therefore prudent to further examine silicone injected cables to ensure that the improvement found elsewhere will be obtained here.

4.3 Research Directions

4.3.1 Accelerated Cable Ageing

In developing research to support asset management of degraded XLPE MV cable networks it is seen that there are two key areas requiring further clarification. First is a need to establish confidence in silicone injection techniques in hot, wet climates and, second, is the need to improve knowledge of the effectiveness of a range of diagnostic techniques in identifying the condition of insulation of cables and other parts of cable circuits such as joints and terminations. In order to examine the silicone treatment it was considered necessary to carry out long-term ageing experiments on real cables from the Ergon network that had been injected with silicone oil. A standard ageing test [20] recommended for new extruded cables requires that cable samples be immersed in water and energised at higher-than-rated 50 Hz voltage with simultaneous thermal cycling by variation of cable current. This a test is designed to produce insulation failures of ten samples of new cables over a period of two years: from the test results it is possible to estimate the parameters of the Weibul failure distribution of the cable.

Figure 3. Test Rig for Comparison of New and Silicone Refurbished 22 kV XLPE Cables

In our research we are currently using the standard test [20] to compare the long-term performance of samples of silicone injected refurbished cable with that of new cables. A view of the experiment is shown in Figure 3. The supply voltage is three times rated phase voltage (38 kV rms). Water is maintained in the baths at a temperature of 50°C and a current of 350 A in the cable produces a maximum operating temperature of 75°C in cable core. Water has a small amount of additional salt to produce free ions which have been shown [21] to permeate into the insulation and affect the growth of water trees. The cable current is cycled between zero and maximum on a daily cycle to simulate the normal variation in cable loading: the variation in temperature assists pumping of moisture into the core and sheath.

4.3.2 Condition Monitoring and Forensic Analysis

The condition of cables in the experiment shown in Figure 3 are being periodically monitored using a variety of methods including pd, FDS, PDC and IRC. The experiment is generating a unique database of condition monitoring information for new and refurbished cables from start of life to end of life. Forensic analysis of failure sites from slices of the cable insulation, some of which is to be dyed and some of which will be examined using sensitive chemical methods, is also expected to provide useful data about the failure modes of the cables. It is anticipated that this project will provide definitive links between the condition of cables as indicated by various tests and the likelihood of failure thereby improving considerably the errors in lifetime prediction that are inherent in the estimation of life from all existing CM techniques.

4.3.3 Monitoring of Cables in the Field

A program of cable condition monitoring is also being undertaken on cables in Ergon's 22 kV network in North Queensland. From these measurements a data base of field CM information is being compiled that will, when combined with the insights provided from the ageing experiment, give an estimation of the condition of the field cables and, their remaining life.

4.3.4 Prioritisation of Cable Repair and Replacement

The improved information about lifetimes of new and refurbished cables from the ageing experiment and the improved knowledge of cable condition from field CM will enable technical assessments of cables to be made and further refinements to the assessments should be possible by considering environmental and network conditions to which particular cables are subjected. However, in order to make decisions based on this information it is necessary to also consider economic factors such as those discussed in section 3. A further economic consideration is that there will be limited capital available annually for management of the degrading cable network. It is therefore necessary to prioritise refurbishment or replacement of affected cable in such a way that disruption to electricity supplies by cable failures are minimised and that the maximum network improvement is obtained overall for funds expended. A computer program that will facilitate the optimisation of the prioritisation task is currently under development.

5 CONCLUSIONS

This paper has described aspects of research currently underway in support of AM of degrading MV cable networks in North Queensland. It is shown that the AM task is complex with key information about the performance of degraded and refurbished cables presently unknown, and the linkage between CM results and cable life not known. The main thrusts of a large engineering experiment currently underway is shown to be aimed at providing unique data on cable life and relationships between CM results and life of a cables in a systematic way on an engineering scale that has not been attempted hitherto.

6 REFERENCES

[1] Steenies EF & Kreuger FH, Water Treeing in Polyethylene Cables (1990) IEEE Trans on Electrical Insuilation, 25(5), 989-1028.

[2] Queensland Competition Authority. (2005) Final Determination - Regulation of Electricity Distribution, Brisbane. 309 pp.

[3] Brennan P. (2003) Asset Management Business Strategies. Proc. of the ESAA Summer School. QUT & Uni. Qld, Brisbane, 15 pp.

[4] Ross R, Smit JJ & Aukema P. (1992) Staining of water trees with methylene blue explained. Proceedings of the 4th International Conference on Conduction and Breakdown in Solid Dielectrics, 22-25 Jun 1992, 456 – 460.

[5] Eager GS, Katz C, Fryszczyn B, Densley J & Bernstein BS. (1997) High voltage VLF testing of power cables IEEE Transactions on Power Delivery, 12(2), 565 -570.

[6] Hoff G & Kranz H-G, (2000) On-site dielectric diagnostics of power cables using the isothermal relaxation current measurements, Power Engineering Society Winter Meeting, 2000. IEEE , 3 , 23-27 Jan, 1593 -1598.

[7] Werelius P, Tharning P, Eriksson R, Holmgren B & Gafvert U. (2001) Dielectric spectroscopy for diagnosis of water tree deterioration in XLPE cables. IEEE Transactions on [see also Electrical Insulation, IEEE Transactions on]
Dielectrics and Electrical Insulation, 8(1), 27 – 42.

[8] Ildstad E & Gubanski SM. (1994) Dielectric properties of water-treed XLPE insulation. Conference on Electrical Insulation and Dielectric Phenomena, IEEE, 1994 Annual Report. 23-26 Oct, 497 -505

[9] Hvidsten S, Ildstad E, Holmgren B & Werelius P. (1998) Correlation between AC breakdown strength and low frequency dielectric loss of water tree aged XLPE cables. IEEE Transactions on Power Delivery. 13(1), 40 – 45.

[10] June-Ho Lee, Sung-Min Cho, Il-Keun Song. (1998) Correlation between the AC breakdown strength and space charge distribution of water tree aged XLPE. Conference on Electrical Insulation and Dielectric Phenomena, 2, 1998.

[11] Hvidsten S, Wrelius P & Christensen, J. (2001) Evaluation of on-site dielectric response methods for non-destructive testing of water treed MV XLPE cables., CIRED. 16th International Conference and Exhibition on Electricity Distribution (IEE Conf. Publ No. 482), 1, 5 pp.

[12] Hoff G & Kranz HG. (2001) Interpretation of dielectric response measurement data from service aged XLPE-cables
ICSD '01. Proceedings of the 2001 IEEE 7th International Conference on Solid Dielectrics. 381 -384.

[13] Gulski E, Wester FJ, Smit JJ, Seitz PN & Turner M. (2000) Advanced partial discharge diagnostic of MV power cable system using oscillating wave test system. IEEE Electrical Insulation Magazine, 16(2), 17-25.

[14] Zhang H, Blackburn TR, Phung BT, Liu Z. (2004) Application of signal processing techniques to on-line partial discharge detection in cables. International Conference on Power System Technology. 2, 21-24 Nov., 1780 – 1785.

[15] Pilling J & Bertini G. (2000) Incorporating CableCure injection into a cost-effective reliability program. IEEE Industry Applications Magazine, 6(5) , 52 -56.

[16] Nannery PR, Tarpey JW, Lacenere JS, Meyer DF & Bertini G. (1989) Extending the service life of 15 kV polyethylene URD cable using silicone liquid. IEEE Transactions on Power Delivery. 4(4), Oct, 1991 – 1996.

[17] Faremo H & Ildstad E. (1996) Diagnosis and restoration of water tree aged XLPE cable materials. Conference Record of the 1996 IEEE International Symposium on Electrical Insulation, 2, 16-19 Jun, 596 -599.

[18] Faremo H. & Ildstad E. (1994) Rehabilitation of water tree aged XLPE cable insulation. Conference Record of the 1994 IEEE International Symposium on Electrical Insulation, 1, 5-8 Jun, 188-192.

[19] Bertini G J & Chatterton WJ. (1994) Dielectric enhancement technology. IEEE Electrical Insulation Magazine, 10(2), Mar/Apr, 17 -22.

[20] IEEE Std 1407-1998. (1999) IEEE Trial-use Guide for Accelerated Ageing Tests for Medium-Voltage Extruded Electric Power Cables Using Water-filled Tanks. Institute of Electrical and Electronic Engineers, New York.

[21] Li H, Crichton BH, Fouracre RA, Given MJ. (1993) Water penetration and initial ion location during water treeing in low density polyethylene. Conference on Electrical Insulation and Dielectric Phenomena, 17-20 Oct, 660-665

Acknowledgement

Insights in to this field provided by a detailed Literature Review completed by Prof. G. Woolsey as part of the QUT/UQ/Ergon project are gratefully acknowledged.

NEW TECHNIQUES FOR ESTIMATING THE CONDITION OF HIGH VOLTAGE POLYMERIC INSULATORS

[a] Birtwhistle D, [b] Cash G, [b] George GA, and [a] Hinde D

[a] *School of Engineering Systems,* [b] *Faculty of Science, Queensland University of Technology,*

George Street, Brisbane, Queensland 4001, Australia.

Abstract: Insulators manufactured from polymeric materials are widely used by many electricity companies although there is concern about how degradation of the polymeric material of insulators might affect their lifetime. The question arises as to how utilities might best manage large populations of degrading composite insulators that are often dispersed over great geographical areas. In this paper the complete life cycle costs of polymeric insulators in large electricity networks are examined and it is concluded that failures of inexpensive components can generate significant real costs to the network owners and great indirect costs to the community. The policy of replace on failure is discussed and it is suggested that the application of condition based maintenance is needed to reduce costs. Failure modes of insulators are reviewed and a new technique for determining numerical indicators of insulator condition, developed by the author and co-researchers, is described that is based on analysis of samples of polymeric insulation taken 'live' from the high voltage network. The condition indicators developed for EPDM and silicone rubber insulation are described and results from sampling in 11 kV and 132 kV networks are presented.

Key Words: Life Cycle Costs, Non-ceramic Insulators, Composite Insulators, Condition Assessment

1 INTRODUCTION

Electricity distribution networks contain large numbers of components, each of which affects the overall performance of the network. Failures of a single small component can often lead to functional failures of major parts of the network with severe consequences. With greater emphasis being placed on financial accountability in the electricity supply industry, it has been recognised as important to examine the real costs of components over their whole lifetime to ensure that informed decisions can be made about the management of the network assets.

Since that late 1980s polymeric insulation has been used in electricity distribution networks at an accelerating rate but polymeric insulation often has much shorter lifetimes than ceramic insulation used in previous generations of equipment. Whilst there are many advantages accruing from the use of polymeric insulation the problems associated with repair and replacement have not been addressed. Many items of equipment are small inexpensive components such as overhead line insulators, surge arresters and bushings. Often these components are replaced on an ad-hoc basis on failure of the previous components and location of the replacement polymer-insulated item is not recorded. There is therefore a situation developing in which equipment with polymeric insulation is being spread in a haphazard way throughout geographically immense areas.

Increasingly, regulators drive network companies to increasing the use of more sophisticated maintenance procedures than the "replace on failure" or periodic maintenance regimes that have been common in distribution networks. There is increasing condition monitoring of components, outcomes from which are used to ensure that maintenance is done only when it is needed. The intention is that condition based maintenance will give lower costs than periodic maintenance with greater reliability than "replacement on failure". The emerging situation with polymer-insulated equipment is not amenable to condition monitoring by existing technologies. In this paper an attempt is made to explore the life cycle costs of large populations of insulators and the progress in the development of new techniques for condition monitoring polymeric outdoor insulation.

2 .ASSET MANAGEMENT OF POLYMERIC VOLTAGE INSULATORS

2.1 Life Cycle Costing Principles

Consider a population of N identical components in a power system. The components may be insulators, line cross arms, connectors, or transformers, or any other large population of network components. In a simple analysis the annualized cost of the component population, C, may be expressed as:

$$C = P + Q + R$$

Where P is an annual interest charge based on the installed cost of the components. Q is the cost of any periodic maintenance and condition assessment made in the year. R is the cost associated with any component failures in the year: this is equal to $Npan$ where p is the probability of component failure in the year under consideration, a is the component cost, n is a dimensionless cost multiplying factor that incorporates the cost of replacement of a component on failure, b, and the costs of the consequences of the component failure, d. We define n as:

$$n = (b + d)/\ a$$

To calculate p we consider the component population to have a mean life L with a standard deviation ΔL based on a normal failure distribution.

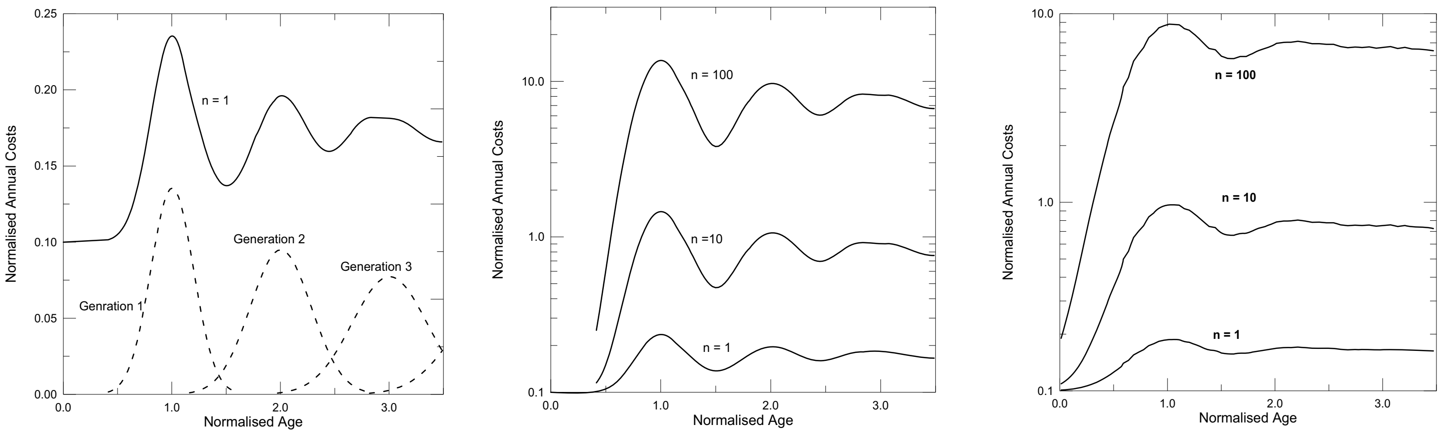


Fig.1(a) $\Delta L/L$ = 0.1 **Fig. 1(b) $\Delta L/L$ = 0.2** **Fig. 1(c) $\Delta L/L$ = 0.3**

Figure 1. Component Costs as a Function of Age of the Asset. Rate of interest 10%. Annual maintenance costs not included.

Normalised annual costs are plotted against normalised age of a population of components in Figure 1. In this figure all costs are normalised by dividing by the initial cost of the population of installed components, ***Na,*** and the age of the components is normalised by dividing the age of the component population by the mean component life, L. We assumed that at year zero the entire component population was installed and that when failures occur components are replaced with identical components.

From Figure 1(a) we see that annual costs are initially constant from year to year and are only interest charges. The costs of the various generations of components are shown as dashed lines in Figures 1(a) and 1(b). These costs are additive to the interest charge costs. Total costs for different values of cost multiplying ratio (n) are included in Figure 1(b). It can be seen from Figure 1(c) that populations with a larger failure spread there is a continuing high cost after the first lifetime due to overlapping of failures of the first and second generation populations.

The above considerations lead to some interesting conclusions about life cycle costing of electricity network equipment.

- Firstly, the life cycle costs of components in a major functional element of a network are inextricably bound up with the relative lifetimes of the components and the major functional element of which they form at part. If components have to be replaced several times over the lifetime of the functional element then life cycle costs need to include the costs of replacing the component possibly several times over the life of the functional element.
- Secondly, the life cycle costs may well be dominated by the costs of replacement of small components.
- Thirdly, it is not possible to estimate life cycle costs of electricity network equipment, unless there is some reasonable information about the life of the component.

2.2 Evaluation of Life Cycle Costs of Polymeric Insulators

In electricity networks it is possible to discern two main types of populations of components. These are: very large populations of rather simple, low cost items (Group A) and smaller populations of expensive items that are often technically complex (Group B). The estimation of life cycle costs of components in Group B is sometimes attempted by distribution network companies. The method adopted is to first estimate the anticipated life of the component and to estimate the total maintenance expenditure based on periodic maintenance carried out at intervals specified by the equipment manufacturer. The life cycle costs appear generally to be determined as the net present value of the purchase price plus the anticipated maintenance expenditure over the life of the component. In the past distribution network companies often had no information about the location or number of Group A components in their networks and the estimation of life cycle costs of components in Group A generally has not been attempted.

In the following section life cycle costs of polymeric insulators are examined. Only costs of purchase, repair and maintenance are considered. The costs of losses and cost of damage to third parties due to failures of the network assets are also not included. All costs are in 2006 Australian dollars and interest or other financial cost multipliers are not considered at this stage.

Insulators in overhead lines may be replaced when they exhibit obvious problems such as cracking, an increase of discharge activity, loss of insulating properties or mechanical failure. Recent examination [1] of some types of MV polymeric insulators, however, has indicated substantial weathering of the insulator sheds of one type of composite insulator after only 3 to 4 years of exposure in coastal locations in the sub-tropics. In this section we will compare the life cycle costs for 11 kV and 132 kV equipment.

In 11kV overhead lines the installed costs of ceramic insulators is about \$20 and composite (polymer) insulators about \$30. There may be about 65 insulators per km in a typical line. Failure of insulators may lead to insulator replacement costs of about \$200. With cost of unserved energy of about \$0.1 per kWh and community costs of unserved energy are between \$2 per kWh and \$10 per kWh, loss of load of 1 MW for two hours leads to costs of \$430 for insulator failure and unserved energy, and community costs of between \$4,000 and \$20,000. Over a 30 year life total life costs of replacing failing insulators with a mean life of ten years would be: \$29,900 per km not including cost of lost load, \$55,900 per km including costs of unserved energy and over M\$ 2.6 per km including the highest values of community costs. In this case the value of ***n*** may be 150 to 680 depending on community costs.

The costs of ceramic insulators for 132kV lines is about \$30 per disc or \$240 for a complete 132 kV string. composite insulator costs about \$400. Typically there are about 9 insulators per km. Repair costs are about \$1200 if a single ceramic disc is replaced by live line methods and about \$6000 per insulator if a composite insulator is replaced. A typical outage time for a composite insulator replacement may be 12 hours. If the line has to be de-energised whilst composite insulator replacement is carried out with consequent loss of load, the cost of unserved energy to say a 30 MW load for a period of 12 hours would be \$3,600, with an additional \$720,000 to \$3,600,000 for community costs. Fortunately loss of this magnitude of load would be rare because of the fact that alternative supplies are generally available.

The above examples are not intended to provide exact life costs. They do, however, provide interesting insights into expenditures in electricity networks. In particular the examples highlight the fact that the repair-on-failure policy can lead to high life costs in MV overhead lines. Life cycle costs will be very high when community costs of lost load are included in calculations and costs are greatly increased if the lifetimes of components are much shorter than the functional element of the network in which they are incorporated.

2.3 The Role of Condition Assessment in Reducing Life Cycle Costs

Traditionally the condition of insulators in distribution network equipment has been assessed from periodic observations from the ground, a technique which provides no information about the condition of polymeric insulators other than about gross degradation such as for example splitting of the plastic housing of the insulator. There is obviously a requirement for some technique to provide numerical indicators of the condition of insulators.

As Group A assets such as insulators are widely distributed over large geographic areas in distribution networks, it is unlikely that measurement of condition indicators of all members of these large component populations will be economic. An alternative approach is to sample the population with particular attention being paid to assets in important locations or in situations where rates of degradation might be expected to be high. If ages of components are recorded, sampling of condition indicators will, over time, enable companies to build up models of the rates at which components age. This information will enable sampling procedures to be refined. Components approaching end of life could be monitored more frequently and replacement programs devised that will avoid failures.

3 CONDITION ASSESSMENT OF POLYMERIC INSULATION

3.1 Failure Modes

Polymeric insulators usually have a resin-impregnated fibre glass core rod for mechanical strength. The core rod is encapsulated by a polymeric housing with weather sheds and protects the structural glass fibre core from the combined effects of moisture, pollution and the electric field. Housings are manufactured from ethylene propylene diene rubber (EPDM) or silicone rubber (SIR) which may also include inert fillers, colouring, stabilisers and UV inhibiters. Commonly insulators of this type called composite or non-ceramic insulators. Composite insulators have two dominant types of failure. The first type of failure occurs when the insulators experience electric flashover (arcing) around the external surface of the housing. The voltage at which such flashovers occur is reduced by the presence of wet pollution on the external surface of the insulator. There is evidence that the surface properties of the insulator play an important role in defining electrical performance during wet and polluted conditions. Aging of the polymer EPDM material can cause a progressive reduction in the surface properties of the material and with it a reduction in electrical performance. Although this reduction may in many cases simply reduce the performance of the EPDM insulator to that of a similar ceramic insulator, it is possible that deterioration can continue further reducing performance and increasing the risk of pollution induced outages. Wet pollution flashovers also occur on SIR insulators but to a lesser extent as low molecular weight oil from the rubber tends to encapsulate surface pollution and maintain surface hydrophobicity which reduces electrical discharges on the insulators surface which are the precursor of flashover. The second type of failure of the housing can occur due to erosion of material caused by intense arcing during wet and polluted conditions or due to degradation of the material leading to the formation of splits in the housing material. The consequences of such failures can be tracking of the glass fibre core leading to mechanical and/or electrical failure. It may be appreciated that both types of failure are preceded by changes in the surface of the housing material and it has been proposed that analysis of small samples of surface material with subsequent physical and chemical analysis of the material will provide numerical indicators of the condition of the housing.

3.2 Condition Indicators

3.2.1 Sampling Methods

Several methods have been developed [2] to remove samples of polymeric material from insulator surfaces. Thin silvers of material are sliced from the insulator surface using a small plane that removes slivers of between 0.25 and 0.5 mm in thickness, width about 10 mm and length not more than about 30 mm. Alternatively traces of the main polymer are removed from the insulator surface by swabbing the EPDM surface with a cotton swab soaked in a suitable solvent. After brief swabbing of the area of insulator of interest, the solvent-soaked swab with minute traces of polymer dissolved in the solvent is returned to a suitable air-tight container from which the solvent is later collected and concentrated for subsequent analysis. In addition special analytical procedures have been developed to analyse the amount of filler in the loose chalky material on the insulator surface. Initially in laboratory experiments, the material was scrapped from the insulator surface. However when the technique was extended to outdoor insulators is was found that wind gusts made the collection of loose material difficult. An improvement in the efficacy of collection of chalky material was affected by collecting the material on the surface of abrasive pads that were rubbed lightly on the insulator surface.

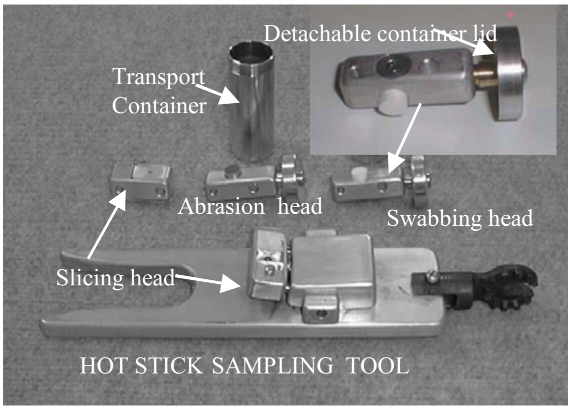


Figure 2. Hot stick tool and fittings for taking samples of shed material

It was not possible, however, to take samples from high voltage lines without de-energising the line. This was seen as a serious limitation to the application of the technique since it has been found [2] that the most degraded parts of insulators tend to be in the high stress regions close to the high voltage ends of insulators. A live line tool was, therefore, developed for taking samples of material from energised insulators. The main components of the live line tool are shown in Figure 2. A pivoted bracket is attached to the main body of the tool. Three types of sampling heads may be clipped to the bracket and secured in position by force due to a permanent magnet. The slicing head is effectively a very small plane that is designed to remove slivers from the insulator surface and capture the slivers in the body of the tool. The swabbing head incorporates a cotton pad that is pre-soaked with solvent. Swabbing heads are stored within a container when not in use are returned to the container after swabbing. The abrasion head is similar to the swabbing head except that the swab is replaced by an abrasion pad. Prior to sampling, and with a sampling head in position on the pivoted bracket, the bracket is moved to a charged position in which a spring is prevented from returning the bracket to its initial position against the main body of the tool by a latch.

Figure 3 275kV insulator sampling

Figure 3 shows the tool in use on a 275 kV insulator. The main body of the tool is inserted along one side of a shed and the bracket with the sampling head attached moves across the other side of the shed. The presence of the shed releases a mechanical latch causing the sampling tool to press against the surface of the shed. Circular movement of the remote end of the hot stick by the operator then causes shed material to be removed by the sampling tool. It has been found that variations of this tool are required for insulators of different designs and a new tool in which the operator exerts a linear pulling force rather than a circular motion has been found to be very effective

3.2.2 Analysis of Material Samples

The sliver surface is examined using SEM to observe the degree of surface cracking of the type shown in Figure 4. A useful numerical indicator of degree of degradation of EPDM is obtained by cleaving the sample sliver of polymer in a plane normal to the surface and making microscopic examination of the cleaved surfaces to obtain a measurement of depth of the degraded surface polymer. A typical result is shown in Figure 5 for a silicone-EPDM blended material. The dark lines in this figure indicate heterogeneous material and suspected de-mixing. X-ray Photoelectron Spectroscopy (XPS) has been found useful in determining the elemental composition of the insulator surface. The indicators have also been obtained from Fourier Transform Infra-red FTIR spectroscopy on polymer weather shed material by ratio-ing peaks in the FTIR emission spectra [1]. The indicators include an oxidation index, a chalking index, and the ester to ketone ratio [3, 4]. The oxidation index is determined as the ratio of spectral peak heights associated with carbonyl (C=O) degradation products (1735 to 1745 cm^{-1}) to hydrocarbon (C-H) at 1460 .cm^{-1}. The numerical indicators give insight into different aspects of EPDM insulation degradation. The oxidation index determines the amount of oxidation. The chalking index quantifies the relative amounts of polymer and filler. The ester to ketone peaks which occur as sub-peaks in the carbnoyl peak have also been ratioed to give a numerical indicator that ratio provides information about the cause of EPDM degradation. It has been suggested [2] that that the limits of the life of the insulating material may occur when:

a. The ratio of the ester carbonyl peak to ketone carbonyl peak determined from the FTIR spectrum is below 0.6.

b. The oxidation index from FTIR is above 0.4.

c. The value of isolated tertiary carbon is equal to or below 0.5.

d. Levels of surface aluminium from XPS are above 7%.

e. The degradation layer is thicker than 20 μm and the width of surface cracks exceeds 7μm.

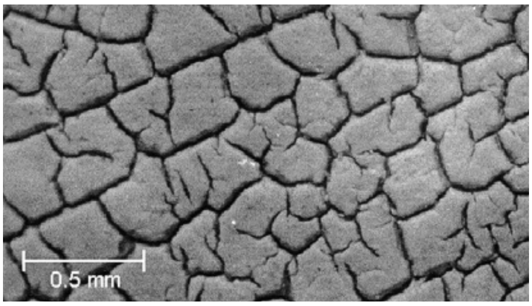


Figure 4 Extensive surface cracking of EPDM insulators from manufacturer C. Top View

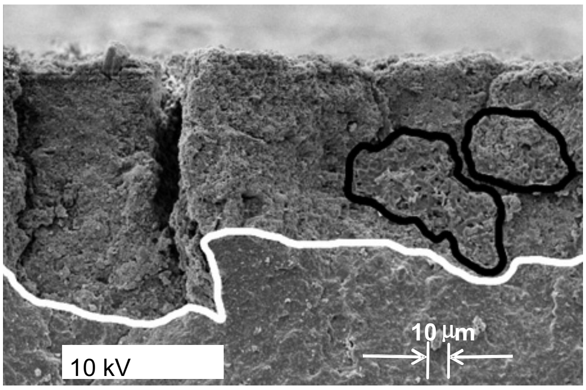


Figure 5 SEM view of the cleave surface of a "White line"- Boundary between degraded and sound

End of life criteria have been difficult to formulate for silicone rubber which is a more complex material than EPDM. However, recent work by Lui et al [5] on the analysis of a badly degraded silicone rubber insulator that was so degraded that some sheds had lost most of the polymer and consisted mainly of the filler gives some indication of the values of indicators at the end of life. Thermogravimetric Analysis (TGA) indicated that there were two major weight losses of from material samples during the TGA heating period, the first due to loss of water from the aluminium trihydride filler and the second decomposition of the polymer. TGA there provides a technique for monitoring the loss of polymer from samples. XPS analysis performed on surface samples of the insulator showed that the worse conditions found were:

1. Total silica was 32 % of total silicone.

2. Atomic concentration ratio C/ Si was 1.71 (whereas for pure PDMS this value was 2)

3. The O/Si ratio increased to 3.5 (compared to 1 in PDMS).

It was suggested that the decrease in weight loss as indicated by TGA and the quantification of silica content from XPS may be the best indicators of silicone rubber insulator life. Acceptable limits will be somewhat lower than the values above. Other indicators based on the measurement of rate of increase of low molecular weight silicone oil on precisely-polluted surfaces have also been investigated recently.

3.3 Application to High Voltage Networks

3.3.1 MV Networks

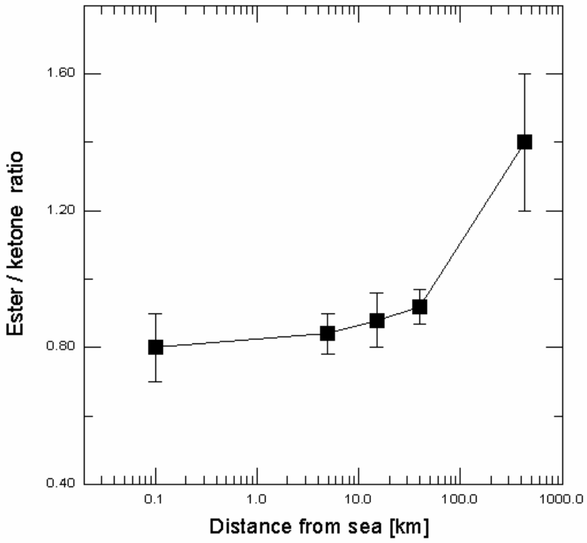


Figure 6. Ester toketone ratio for insulators as a function of distance from sea. (Manufacturer A)

In medium voltage (MV) (less than 66 kV) electricity networks most currently installed insulators are EPDM with a smaller proportion of EPDM/silicone rubber blends. The performance of polymer insulators in tropical climates is, however, not well known. An investigation into condition assessment of EPDM insulators in north-east Australia commenced in 1996. Six participating utilities supplied almost fifty insulators produced by seven manufacturers. The insulators came from locations characterised by different climatic conditions. These included the desert-like climate of inland Australia with daily temperatures of up to 45°C, low rainfall (600 mm/year) and low humidity (30%) and the humid tropical climate of coastal regions with abundant rain (up to 2000 mm/year), high humidity (80%) and temperatures up to 30°C. Australia also has one of the highest levels of ultraviolet exposure in the world. The total annual dose of ultraviolet radiation is more than double than that experienced in Europe [6] Indicators were determined for insulators from three manufacturers. Sheds of insulators from manufacturers *A* and *C* were made from a blend of EPDM and silicone rubber. Sheds of insulators from manufacturer *B* were made from EPDM. Three different types of location were recognised: sea, coast and inland. The sea side locations refer to areas in close proximity to the sea – not more than 200 metres and with direct exposure to salt spray. The coastal locations are more than 200 metres but less than 100 km from the sea. The inland locations are more than 100 km from the sea. Up to six insulators per location were analysed. The extent of the degradation of the insulators from manufacturer *C* was reflected in high levels of oxidation index (>1.5) and chalking index (up to 4). The depth of the surface degradation layer was approximately 100 µm and some of the surface cracks were deeper than 50 µm. The depth of the surface degradation layer of insulators from manufacturers *A* and *B* did not exceed 30 µm. The heterogeneous nature of the degraded surface layer of insulators produced by manufacturer *C* indicated that poor mixing between EPDM and silicone rubber may have taken place, see Figure 4. It was found that the ester to ketone ratio was significantly higher (>1) for inland insulators (Roma, Toowoomba) than for insulators at the sea or in coastal locations (<1), see Figure 6.

The oxidation index for the sea side insulators of manufacturer *A* was greater than for the coastal and inland insulators of the identical manufacturer. The only exception was an insulator at in Goodna that was installed close to the Brisbane river in an area suffering from frequent fogs in the cooler months. Both, salt spray at the sea, and fogs in the river area caused extensive dry-band arcing which damaged insulator surface and resulted in the high values of oxidation index. The oxidation index for the sea-side insulators of manufacturer *B* was much smaller than those of manufacturer *A* even though the insulators of both manufacturers were installed on the same pole. However, insulators of manufacturer *B* showed more extensive surface damage than insulators of manufacturer *A*. The latter point is confirmed by the thickness of surface degradation layer which is approximately 10 µm for insulators of manufacturer *A* and 20 µ m for insulators of manufacturer *B*. Insulators from manufacturer *C* performed worst. They had uniform white chalking on all surfaces and showed extensive surface cracking. It has been hypothesised that solar ultraviolet radiation can cause photo-oxidation of polymers leading to the formation of ester-type degradation products. When meteorological data from the regions were compared, it was found that the inland location (Roma) has 150 clear and 70 cloudy days a year. The coastal location (Beenleigh) has 80 clear and 120 cloudy days a year. As clouds act as a partial screen for solar ultraviolet radiation and both the inland and coastal locations are at approximately the same latitude, the inland insulators received higher doses of ultraviolet radiation and this resulted in higher proportion of ester-type oxidation products.

It has also been shown that thermal oxidation of polymers similar to EPDM can favour ketone oxidation products [7]. In coastal areas, high humidity and salt spray may lead to increased dry-band arcing with subsequent thermal oxidation of the polymer and thus decrease in the ester to ketone ratio. A similar phenomenon has been observed in results of tests from a 275 kV EPDM insulator at a coastal location of similar latitude to Roma or Beenleigh. The ketone to ester ratio was greater at the HV end than at the ground end. This might indicate more dry-band arcing in the high field region near the HV conductor. Annual rainfall might also have played a role in the phenomenon observed in Figure 6. The inland location, Roma, has a lower average annual rainfall (600 mm/year) than the coastal locations, Beenleigh (1000 mm/year). Water soluble oxidation products might, therefore, have been washed away by rain.

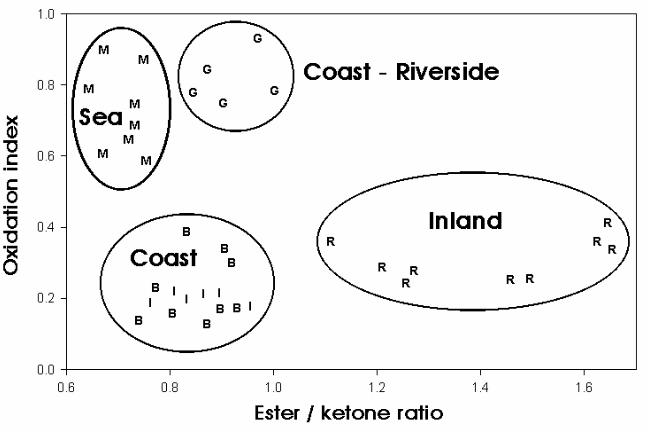


Figure 7 Scatter plot relating the oxidation to the ester toketone ratio for insulators from manufacturer A.

B-Beenleigh, 15 Km from sea, 6 Insulators

G-Goodna, 40 Km from sea, 1 Insulator

I-Ingham, 5 Km from sea, 1 Insulator

M-Miami, 0.1 Km from sea, 3 Insulators

R-Roma, 425 Km from sea, 2Insulators

Effectiveness of the condition is demonstrated in Figure 7, where a two-dimensional scatter plot relating the oxidation index to the ester to ketone ratio for insulator samples from manufacturer *A* is shown. Several distinct groups in the data can be identified. The sea-side insulators from Miami and the coastal insulators from a foggy area in Goodna have high values of the oxidation index due to dry-band arcing, though the coastal insulators have slightly higher ester to ketone ratio due to more extensive UV damage. Inland insulators from Roma with prevailing sunny weather and intense UV radiation have the highest values of the ester to ketone ratio. Coastal insulators from Beenleigh and Ingham form an overlapping group with low values of the oxidation index and the ester to ketone ratio.

3.3.2 High Voltage Insulators at End of Life

During the 1980s a 132kV line was erected in a coastal location (over 1 km but less than 20km from the sea) in South East Queensland. At that time cantilever post insulators having weather sheds manufactured from EPDM were used. The insulators used came from two different manufacturers. The most common type consisted of 16 individual EPDM weather sheds approximately 230mm in diameter fitted on a glass-fibre reinforced resin core, sealed with silicone grease and clamped between two galvanised end fittings. The other consisted of 22 sheds of 210mm diameter. After 10-12 years service a small number were removed due to longitudinal splits in the rubber in the core adjacent to the HV end. Visual examination revealed that significant surface changes had occurred. This included extensive chalking of all polymer surfaces and the development of a rough surface in place of the original smooth surface. The longitudinal splits in the polymer covering the core at the high voltage end had exposed the glass fibre core to the ingress of water and dust. The rubber in this area had a “wrinkled” appearance due to ageing of the EPDM.

The longitudinal splitting of the core near the high voltage end indicated a loss of elasticity of the EPDM. This is probably a result of cross-linking during oxidation which makes the material more rigid. There was also some evidence from chemical analysis of the oxidation products recovered by swabbing the surface that there had been splitting or scission of the polymer chains during oxidation. This is further evidence of extensive thermal oxidation of the EPDM.

The apparent extensive surface “chalking” observed was quantified using diffuse reflectance FTIR. The analysis of selected samples by X-ray Photoelectron Spectroscopy (XPS) confirmed that the atomic composition of aluminium on the surface was between 8 and 10%. When new specimens of EPDM insulator material are examined by XPS, no aluminium was seen on the surface. Scanning electron microscopy (SEM) of the surface revealed a layer of degradation of at least 30µm as well as wide spread surface cracking. Towards the bottom of some of the broadest cracks stretching of elastomer can be observed. This is consistent with loss of elasticity perhaps caused by oxidative crosslinking of the polymer as mentioned above. It is believed that this small scale cracking eventually led to the complete splitting of the EPDM in the core of some insulators.

4 LIFE CYCLE MANAGEMENT OF POPULATIONS OF COMPOSITE INSULATORS

The provision of a series of field-aged insulators by the local utilities has permitted the development of ‘end of life’ criteria in terms of numerical condition indicators for EPDM insulators and formulation of tentative criteria for SIR insulators. If a quick assessment of a number of specific EPDM insulators is required, a single chemical test - surface swab followed by FTIR emission spectroscopy - can indicate insulators that may be “at risk”. Such insulators can then be studied further by both chemical analysis and SEM, and a more reliable indication of condition obtained. Utilities are then in a better position to make decisions about the reliability of the insulators and selective replacement can be made prior to serious failures occurring.

EPDM Insulators from a line indicating high levels of oxidation but a high ester to ketone ratio suggest that although the insulators may be heavily degraded, the combination of the insulators surface condition and pollution severity may be such that significant surface discharging does not occur. The techniques presented provide valuable information for managing both types of failures. The combined measures of the oxidation index and the ester to ketone ratio are promising because they appear to provide indication of both the degree of deterioration and the presence or absence of electrical discharge activity. For example insulators from a line indicating low ester to ketone ratio and high oxidation would indicate that these insulators are degraded and have experienced high levels of surface discharge activity. This in itself could indicate that the combination of pollution severity and material condition is such that there is an increased risk of pollution flashover. By considering the environment, and the importance of the line a decision could be made whether further surveillance such as leakage current monitoring is required. The indication that high levels of surface discharging have occurred may also indicate that erosion type defects are more likely to develop. This could trigger a program of visual inspections or assessment using some of the more advanced techniques such as infrared scanning, or electric field measurements. Insulators that show a low ester to ketone ratio but have a low oxidation index indicate that arcing has occurred, but this has not yet produced significant deterioration. It is suggested that although these insulators may still be in good condition, they may be in an environment where they are likely to deteriorate and should be monitored.

The usefulness of the CM method described in this paper lies in the fact that it can be used to sample the condition of the insulation material from time to time to determine the numerical indicators of the condition of insulators that are part of a larger population. In the initial sampling it should not be necessary to sample large numbers of insulators but rather the samples should be restricted to areas where environmental conditions are severe. If sampling indicates that there only small changes in the indicators over time then it may not be necessary to make other types of inspections that require line patrols. If the indicators of shed condition do, however, indicate that there has been substantial degradation in the properties of the shed material then it will be necessary to make detailed assessments of the insulators by techniques that have been developed to locate defects. It will be a simple matter in sampling to identify particular types of insulators likely to experience early failures and to put in place programs of replacement at an early stage. To enable sampling programs to be effectively implemented electricity companies will need to keep records of where insulators of particular types and manufacturer are installed: this is currently not done for any but the highest voltage types.

5 CONCLUSIONS

Composite insulators are widely used in electricity networks but their lifetimes are less than that of previous generations of insulators. This may lead to failures of overhead lines that can cause loss of supply which it has been demonstrated can be expensive for utilities and very expensive for the community over the life of the asset (such as overhead lines) in which insulators are used. A new type of technique for monitoring the sheds of composite insulators has been reviewed and it has been shown that it is possible to take very small samples of insulator shed material "live line" and to analyse the samples by a number of technique that provide numerical indicators of insulation condition. These techniques are well developed for EPDM but are still being developed for SIR. The techniques have been applied to high voltage networks throughout Queensland.

6 REFERENCES

[1] Blackmore P, Birtwhistle D, Cash G & George G. (1997) In-situ condition monitoring of EPDM composite insulators using Fourier transform infra-red spectroscopy. Proc. of the 5th International Conference on Properties and Applications of Dielectric Materials, 2, 910 – 913.

[2] Birtwhistle D, Blackmore P, Cash G, Gillespie T, and Krivda. A, (2000) Application of advanced chemical analysis techniques to the assessment of composite insulator condition, CIGRE Paris.

[3] Birtwhistle D, Blackmore P, Krivda A, Cash, G & George G. (1999) Monitoring the condition of insulator shed materials in overhead distribution networks, IEEE Trans on Electrical Insulation, 6(5), 612-619.

[4] Blackmore P, Birtwhistle D, Cash G & George GA, (1998) Condition Assessment of EPDM Composite Insulators using FTIR Spectroscopy, IEEE Trans. on Dielectrics and Electrical Insulation, .5, 132-141.

[5] Lui H, Cash G, Birtwhistle D & George G. (2005) Characterization of a severely degraded silicone elastomer HV insulator-an aid to development of lifetime assessment techniques. IEEE Trans on Dielectrics and Electrical Insulation, 12(30), June, 478 – 486.

[6] A. Davis and D. Sims, (1983) *Weathering of Polymers*, Applied Science Publishers.

[7] George GA, Celina M, Vassallo AM & Cole-Clarke PA. (1995) Real-time Analysis of the Thermal Oxidation of Polyolefins by FTIR Emission, Polymer Degradation and Stability, 48, 199-210.

MANAGING THE INTANGIBLE ASSET

R. B. Ward

Faculty of Engineering, University of Hew South Wales, Sydney, New South Wales

Abstract: About fifty years ago a superintendent of apprentices taught his very junior people that management involves the use of four resources: men, machines, materials and methods [1]. (More recently this author added "money" to those four, but that is another story.) All four are assets and require maintenance, one way or another, and while much has been written about maintaining machines, and, indeed, methods, there seems to be little about maintaining people. The paper following from this abstract shows why "people-maintenance" is needed, examines some of the difficulties (such as the wide range of levels in an engineering organisation), shows the risks in doing, or not doing, this form of asset management, suggests how the asset may be condition-monitored and what standards may be used, and points out the cost of good people-maintenance (and the cost of not maintaining the people asset).

Key Words: management, people

1 INTRODUCTION

The "intangible asset" in the title is people, who are, actually, of course, quite tangible in the sense they exist and can be seen and touched. However, even though that's recognised as being obvious there's a definite impression that people aren't as "tangibly real" as other physical assets such as land, buildings, machinery and similar asset items.

Before progressing to a discussion on why people should be "maintained" we need to state why this is of interest and concern to engineers. That's evident from the list of assets mentioned in the above summary, men, machines, materials and methods. The first three are certainly "physical", the fourth encompasses all types of spoken and written instructions, covered in today's language by the term "intellectual property", perhaps not as "physically-real' as the others (unless existing as documents), but still of value, hence an asset. So if people are one of the assets engineers use and maintain, we submit that they, like the others, should be maintained. People need maintenance just as the other assets do.

So, to the connection between engineers and people. Engineers are familiar with the use of all four of those items, most familiar with the maintenance of machines, materials and methods, all of which need engineering work to overcome wear and tear of machine parts, selection of improved materials, and adjustment and correction of methods as these require updating. Engineers seem to be least familiar with using men, or, to be more general, with people, yet people are a significant asset in any business, even one in which machines and automation have taken over the physically demanding aspects of work. The reason for this appears to be the general task-orientation of most engineers, implied by Samson [2] and indicated by the results of this author's experiments, using management case studies with senior engineering students. With rare exceptions the students primary concentration was on technical factors, then on management issues, and only very occasionally commented on what was presented between the characters.

2 WHAT CAN WE FIND ABOUT PEOPLE-MAINTENANCE?

Maintenance in a purely engineering context is, of course, principally about repair of machines and related equipment with some development in addition, and we might expect maintenance of people similarly to be about repair of wear and tear, but only partly, it should also be about development and support. All contain difficulties; as an illustration of development problems, in a novelised version of a case study sequence this author has presented discussion of what's facing managers in a factory, and one character described the situation, intended to develop those involved, as a test to destruction. The other, the person with an engineering background, pointed out that such a test should not be performed in service, only under laboratory conditions.

Making that statement on the content of maintenance leads into seeking literature support for what the management gurus may have given us on people-maintenance. Consider, first: why "maintain" people? For the organisation's benefit? Or to help the individual person? Or both? That last alternative is the most sensible, provided it truly covers both, and we might expect some reference to this topic to appear in the many texts on management, for surely people-maintenance is part of the general management function. However, search through a considerable range of books, on many aspects of management, has shown this topic is not just remarkably thin but effectively non-existent. Yes, much has been written on how to manage people, but the motive behind such explanations and directions comes out as being for the benefit of the organisation, or for its management, not for any benefit of the actual people being managed.

As an example, going back nearly fifty years, there's a prime example of that omission in Herzberg's presentation of motivation and hygiene factors, both of which contribute to motivating people [3]. The first are positive factors which, *if present*, will *motivate* the individual, result in high performance and cause the individual to experience job satisfaction, examples being the work itself (doing something one likes), responsibility, achievement, recognition, personal growth. A very simple example of the recognition factor is being addressed by name by someone way up the management structure from oneself; which gives even the most cynical person something of a lift. The second, hygiene, or *dis-satisfaction avoidance,* factors, are the opposite, if these are *lacking* they *de-motivate* the person, but providing them will *not* provide a source of motivation. They must be present, adequately, or dis-satisfaction will arise. Examples of these are salary (in the sense of adequacy or improvement), relationships with those above and below, satisfactory work conditions, status in the organisation, security. The most curious item in this is that more money does not motivate people; a seemingly unlikely concept, but it came out of Herzberg's published research, if adequate it only avoids dis-satisfaction.

So that's what he found, what he recommended managers should apply. But why? The reasoning was simply to motivate workers to work better, not for their own benefit but (as already remarked above) for the benefit of the employing organisation. But what about the positive motivation factors contributing to job satisfaction? Surely that benefits the worker? True, but the reason for improving job satisfaction, in this context, is to have the worker work better, for the organisation's benefit.

As another example, take Reason's work on human error [4], which covers the various forms of error, much of what causes them, and (to some extent, and with other writers) how errors can be prevented from occurring. Simply, forms of people-maintenance can be applied to reduce error occurrence. And why should errors be prevented from occurring? Because they're bad for the employing organisation.

A final example: an Australian management text [5] has a whole chapter on Human Resources Management, covering selection, recruitment, selection, training, etc. There's discussion of specialisation of work and the writers point out that this movement tends to make work less interesting, then recommend the use of "Job Enrichment" as a way to overcome employees becoming automatons. However, the reason for so doing is more to improve company results than to perform people-maintenance. Echoing Herzberg.

This author has therefore been able only to conclude there's little, if anything, in the management literature on people-maintenance, so the author's thought direction turned towards observed examples in which nothing was done to support people in the sense of maintaining their abilities, or something wrong was done to cause harm in some manner, and to suggest from these what could, indeed should, have been done. Like anyone with experience as an academic this author shudders at the use of anecdotal evidence but assures the reader these are real examples, from the industrial experience of several engineers and managers. None are fictional.

3 WHY IS PEOPLE-MAINTENANCE NEEDED?

Before progressing to those examples we need to look at the reason for management going to the trouble of exercising people-maintenance.

The most obvious reason relates to the health of the employed people. People get sick and insist on carrying on because "they're indispensible". People work excessively long hours because they want to get the job done. People defer taking annual leave because "there's something needs to be finished", and sometimes continue to defer so that there's a series of "somethings" which roll on one after another.

As well as disturbing the individual's life all those activities (or lack of action) have counter-acting factors in a well-managed firm. Department heads are known to shoo sick people back home, partly for their own sake and partly to protect the rest of the office. Department heads have been known to curtail long weekday hours, and weekend work, not only to limit costs but because people need at least a short break after a certain number of days with the back to the wall and the nose to the grindstone. Department heads have also been known to insist on annual leave being taken for the same reason, people need a longer break after months at work, and many firms have limits to how long a person can put off taking annual leave.

This author has had experience with the result of both no weekend break and taking annual leave. The first was a need to work through three weekends to finish a project on time, he admits at the end of the last weekend he had completely lost track of the days of the week, and slept for well over twelve hours when the work was completed. The annual leave example was much happier, a machine failure occurred while he was absent on leave, the department went through two weeks of agony trying to solve the problem but could only keep the system staggering along. When he returned, fresh from two weeks relaxation, he was able to point to the offending component after a couple of hours thought, but through the years since he has admitted to "cheating", seeing the answer easily by not being bothered by the dreadful two weeks experienced by others.

Keeping employees physically healthy is important because having a person absent costs something, easy to determine when a production person is missing and another on overtime or casual rates replaces, perhaps hard to measure in the case of staff people, but it's a finite amount. What may be termed mental health is much harder to see, it's related to having people satisfied, even preferably happy, in their jobs, which may require management playing confidence tricks on some persons.

On-the-job mental health, related to job satisfaction, which is also related to the development aspect of people-maintenance, is much harder to explain. In an organisation some people have no wish to move up through the ranks (this author has known some), but those who do so wish may find not everyone can move upwards through the hierarchy because there's a numbers limit at every level. So, for those who have some form of upwards-seeking drive means must be found to satisfy that if they're to be retained. Extra education may serve, miscellaneous extra duties, peripheral tasks and projects may be found in reasonably large firms but are usually hard to find in smaller companies.

So, now to the observed and reported examples.

4 EXAMPLES OF POOR PEOPLE-MAINTENANCE

The case of N brings to mind the aphorism about the destination of the road which is said to be paved with good intentions. At the time of his leaving the employer N had been there nearly twenty years after graduation, had moved through several departments, all of which had passed him on to another with reports generally of the "he's all right, but -- - " faint praise nature, none wishing to keep him. He was finally moved into the engineering department, which covered a mixture of maintenance and development, and after a certain amount of fumbling over where he should fit he was put in charge of maintenance at the principal site. During one very wet weekend equipment in part of a production system was flooded, causing lost output, N was notified, didn't act promptly, and on the Monday said quite openly: "It's not my responsibility," referring, of course, to getting the repairs done expeditiously. He was quietly but firmly dismissed.

One may wonder how someone with his background could have said that, considering he had been given, and had accepted, the task of overseeing maintenance of the site. The only answer seems to be a lack, through the twenty-or-so years, of his learning that supervisors, as lower-level-managers, have responsibility for, and are accountable for, effective results in the work to which they are assigned, which attitude could have been caused by the concentration of power held by the more senior managers under whom he worked.

There's a completely different situation in the case of R, who was in charge of maintenance in a small chemical firm through several years, during which R had become almost a personal friend of the Production Director, one of the firm's owners. The curious feature of this situation was R was neither an engineer nor a tradesman but had a more-or-less 'handyman' background. He had been with the firm since very early days, had seen much of the process development, was familiar with the relatively primitive technology used there, was able to perform any elementary maintenance jobs himself, and had access to some outside contractors for specialised work. And, of course, he was friendly with one of the two owners.

The time came when the firm took on a more advanced technical system, decided it was time to employ a professional engineer, went through the usual procedures, and appointed one. Although R showed no overt antagonism towards the engineer his resentment was evident to outsiders, and increased when a qualified tradesman fitter was added to the 'Engineering Department', all understandable from a human behaviour viewpoint because R must have felt 'put down' by the changes; he had moved from being the one solely in charge to being at the bottom of the pecking order.

With the undoubted benefit of hindsight it's evident that management's people-maintenance of R took a wrong turn years earlier, by advancing him in position without ensuring he was gaining appropriate skills to match his position. Or by keeping him at a level appropriate to his knowledge-base, which of course included his extensive background of the early plant development. Either way would have been difficult, but the result of the situation was fairly inevitable, in time the fitter and the engineer both left, leaving R in charge again.

There are similarities between the above and the case of H, but also some differences. H had been employed as electrician in a medium-size firm, in a small engineering department headed by a professional engineer. A time came when the engineer had left, and H was promoted (probably by a "by default" process) to the one in charge of engineering. Some years later the firm was swallowed by a larger competitor, with H remaining in charge of engineering at what was now the secondary site of the overall fairly large firm, but he was not made responsible to the engineer at the primary site. Then the firm appointed another engineer who would manage maintenance and projects at both sites, meaning H would report to him, all very well, but complicated by the engineer having worked in the earlier small firm, over ten years earlier and before the amalgamation, in a position more or less level with H's original position There was no overt friction between them, but H appeared to resent having to follow more formal procedures his new manager required.

During a maintenance shutdown H supervised replacing the packed gland on a large-diameter shaft with a mechanical seal, and in the first assembly one of the two carbon components broke. The parts were stripped and a second carbon was fitted: it also broke. The third assembly also broke a carbon, the last spare available, so the packed gland was replaced. H immediately handed in his resignation, could not be persuaded to withdraw it (persuasion from his manager and many friends on the site), and he left. A few years later his manager met him and found he was much more relaxed, stress-free, and enjoying work as an electrician in a local bus depot.

The similarity between the R and H cases is in the advancement of an unsuitable person in an organisation to a level without ensuring the person has the necessary features. The principal difference is the first was resolved in favour of the person (though inappropriately), the second was resolved with short-term discomfort but for the long-term benefit of both the person and the organisation.

Another different example comes from B's history with a very large firm, in which B provided a specialised service to production plants in several states for many years. He was highly respected for the extent of his knowledge and his ability to provide answers for problems, and he was well liked as a person. There was only one problem: he had no successor. Several attempts had been made to provide one by appointing a junior person with appropriate education (in the same specialised area) as his assistant but the associations only lasted a few months after which each assistant left. There appeared, in this, to be nothing like the situation in which the emperor (or king, or baron) who fears the crown prince (or other person) will stage a palace coup and depose him - - - a state of affairs which occurs as often in industry and commerce as in medieval, or even some modern, royalty - - - the problem was simply that B was so truly competent in his speciality he couldn't hand over work items for another to execute, he was the epitome of "hands-on".

The *word* (a form of communication stronger than gossip) which went around the production plants was that B was a nice fellow but impossible to work with, the short-term assistants said enough to make that understood. Obviously, B had no idea of why an assistant should be trained to provide succession or how an assistant could be trained, and also seemed unable to realise a well-trained assistant would make B's life easier. His inability to accept the reasons for training and developing another indicated poor people-maintenance on his part, and the way in which the company (personified by the general manager above him) tolerated his behaviour indicated equally poor people-maintenance within the organisation.

The word 'gossip', used above, relates to the case in which A was involved. A was a production supervisor in a Sydney company, reasonably competent in that role, but extremely, indeed exceedingly, competent as one who supplied his manager with fragments of information from within the factory. Anyone who has worked in a factory, particularly one which runs a rotating shift system, knows that there's a certain amount of sheer bitchiness (for want of a better word) between people, between shifts, between workers and management, and this grapevine of communication forms a lubricant by venting irritations caused by rotating shifts. (That mixture of metaphors seems to describe well what happens.) The rotating shift system disturbs life in a very real manner and there has to be a way of getting rid of the irritation, so a sensible management may listen to what's noised around in the factory but ignore most of it.

In this case, however, A was encouraged as an open and active information conduit from workforce to management. The information was often used against individuals, as one example he passed on a remark which he overheard, made in confidence from one person to another, and the speaker was given what might be termed verbal chastisement. There is no doubt A was, as a person, maintained very well in the function he performed, but keeping a person in such a role is poor people-maintenance for two reasons: it gave A an exaggerated impression of his value and it reduced the workforce's opinion of the management.

The evidence of poor people-maintenance in J's case showed by how he was managed and how he managed those under him. Both were exposed after he requested an additional professional engineer for the project he was leading as employee of a major firm, and in response the company's human resource office hired an engineer as a casual-contractor through an agency. Without, it seems, in retrospect, asking sufficient questions. With a very few weeks the casual employee had performed very little work, had observed that others in the project group were also effectively inactive although the project manager was always quite busy with phone calls, seeing other contractors, or out meeting others, and the project manager seemed to place more importance on having so many people around him than on having them all engaged. The casual employee went to the human resources office, reported he wasn't being effectively used and outlined his observations to a senior person, who admitted the office had no knowledge of how members of the project group were being utilised, stated the situation would be checked and necessary action would occur. The casual employee was shifted to other work via the agency. The project group carried on without him.

Summing up on J's case: the section of the firm's management hadn't monitored the behaviour of the project group and therefore weren't able to apply appropriate maintenance, and the manager of the project group may have been aware of what was happening but wasn't controlling its action as he should have.

5 WHAT WENT WRONG?

The above examples, of what has been heard to have gone wrong in managing people, show, more than anything, that the seeds of organisation troubles originate more often from the behaviour of senior people, although the fruits show up lower down. Summing up, we have:

N: his unsatisfactory behaviour was noted by department after department, but nothing was done to repair that deficiency, which finally came out by showing a lack of understanding responsibility and accountability.

R: the firm's part owner allowed his friendship to show in the workplace, which at least contributed to R's attitude of "owning" the engineering function.

H: the management promoted H to supervising the site's engineering department because "he happened to be there", not based on his assessed suitability; in his favour we point out he recognised he wasn't suitable when a minor disaster occurred.

B: his work was more than satisfactory and he enjoyed what he was doing, so both he and the management were happy to allow him to continue as a one-person department.

A: his competence as a production supervisor and as an informer cannot be questioned, but for his manager to encourage the latter and use the information raises questions of organisational ethics.

J: there appears to be only two possible reasons for J's behaviour, either he was simply empire-building and wanted to be surrounded by underlings, or he really believed more people were needed for his project.

6 WHAT RESULTS FOLLOWED?

Summing up, now, the results of the above examples of poor people-maintenance:

N: he was dismissed and had to find other work.

R: the firm went through a disturbed period of covert in-fighting then loss of experienced people.

H: he left the firm where he'd worked for nearly twenty years, and found happy employment elsewhere.

B: he carried on in the same manner to retirement; what happened after that is unknown.

A: the management continued to use him, which led to some alienation with the workforce and some staff.

J: the casual employee continued employment elsewhere, J finished the project and was retained.

7 WHAT SHOULD MANAGEMENT HAVE DONE?

Hindsight, as remarked above, can produce wonderful answers, all of little use after the horse has bolted through the stable door, and here's an *oh-so-obvious* review of what should have been done to apply people-maintenance to the persons in the illustrative cases given.

N: a cultural change should have been introduced, to bring in acceptance of responsibility and accountability.

R: the friendship between ranks should not have been apparent, and used, in the workplace.

H: a period of understudying a senior person should have been provided, if his position were to be confirmed.

B: management should have insisted that B would accept, and work with, an assistant.

A: the management culture should have been changed to stop the use of gossip as a disciplinary tool.

J: the human resources department should have questioned the request, and monitored the use of staff.

8 WHAT WERE THE COSTS?

The costs of cases such as the above are usually shared by both the person and the firm, and although numbers cannot be quantified the costs can be indicated.

N: Personal: loss of several weeks income, loss of personal confidence. The firm: loss of an experienced person, cost a hiring a replacement (who, actually, was a dead loss).

R: Personal: R lost nothing, indeed, he probably gained by carrying on after the others left (they had costs). The firm: loss of experienced people, cost of hiring replacements.

H: Personal: H lost some income and was unhappy for a few weeks, but actually benefited, long-term. The firm: no cost, existing people were shuffled around to cover the gap left by H leaving.

B: Personal: no apparent cost, but behind the scenes one can conclude he was working harder than needed. The firm: no apparent cost until B retired, then his expertise had to be replaced (result not known).

A: Personal: A continued in his dual position, with no cost to him apart from general dislike by workers. The firm: no apparent cost, but the workforce became recalcitrant and A had to be moved to another site.

J: Personal: not known, but one can only assume human resources did question J's management.

The firm: the project costs *versus* budget are not known, but that firm was inclined to be liberal.

Having stated, above, what can be excavated from decades-old history, what can we summarise? In every case the poor or inadequate people-maintenance caused a cost to someone, sometimes the person who was poorly managed, sometimes others around that person. In most cases there was a cost also to the employing firm; replacing employees takes time and uses the services of people who have other work to do.

9 A FURTHER EXAMPLE

This example has been left to near the end of this paper for several reasons which include it being not from industry and its providing another, different, demonstration of the cost angle.

An academic with a considerable number of years service in a university retired after accumulating an extensive record of publications and supporting scholarships. About a year after retiring one of his children died in an accident and although the faculty knew of that no condolence messages came to the family from anyone in that faculty.

The accreditation of further publications was transferred to another university, about four papers per year, meaning the related funding from Canberra would be received by that university, which also received the private scholarship funding.

A small example, but it illustrates how lack of "doing the right thing", which is what people-maintenance is all about, can impact on an organisation, in this case by loss of funding.

10 SOME EXAMPLES OF GOOD PEOPLE-MAINTENANCE

After the above series of managers who seemed to have no concept of how to maintain their people we can gain relief by reviewing a case or two in which good people-maintenance occurred.

The subject of this case, P, was a junior project engineer in a medium-sized company, still studying for the TAFE certificate, and, recognised in retrospect, given work beyond his age (early twenties) and experience (this was his first job of this type, and he had not had opportunity to understudy someone senior). The senior engineer to whom this junior person responded has told how the junior fell behind with a critical project but for a few months reported satisfactory progress, too embarrassed to admit he was running late and hoping he'd catch up. When the actual, and lingering, progress was discovered the senior succeeded in getting the fabrication step completed very quickly, and had the junior supervise the installation stage to completion over a weekend.

The senior engineer has told this author that he talked the situation over with the junior, who expected to be dismissed. But through the busy weekend the senior had mentally reviewed some of his own stupid acts through the previous twenty-or-so years and all he felt he could ethically do was to deliver a strong lecture on what had been done wrong and 'don't ever do that again' - - - particularly as the junior had put in many weekend hours finishing the work.

This present author was involved in a good people-maintenance experience, in the dark ages over thirty years ago, when a female university engineering student, K, asked for vacation experience work. His first reaction was: "No, that's not possible!" with an immediate afterthought: "Why not?" The student was hired for six weeks, did some useful work, gained on-the-job understanding of some of her learning, learned some of the art of supervising when she had to exercised her voice by screaming at a fitter who wouldn't do what she told him to do, wrote a satisfactory report, and went off to graduation (and marriage).

11 DISCUSSION OF THE GOOD PEOPLE-MAINTENANCE EXAMPLES

Having thoroughly analysed the results etc of poor people-maintenance, some similar comments are needed on the above two examples of good people-maintenance.

The ethical dilemma which appears to have faced the senior engineer in the first example was how to deal with the junior in a manner delivering justice, when he remembered he'd been caught out under similar circumstances as a junior but had been given opportunities to correct what had happened. As with most ethical dilemmas this had no truly right answer, so he acted as he did, hoping the result would be worth his effort.

Concerning the junior, his attitude became much more focussed on his work, still able to share a joke but more intensely purposeful, with further proof that was good people-maintenance showing about two decades later, when the junior, now older of course, was found to be Pacific Area Manager for a large international company. Concerning the firm in which the incident occurred, nothing was lost, and the work was completed on time. The senior engineer experienced the satisfaction of solving the work-related and ethical problems.

In the second example everyone gained, the student got the work experience needed to satisfy the graduation requirements, the firm gained by having an extra supervisor during a maintenance period and by obtaining the information on the reports, and the author gained by the experience, rare in those days, of having a female colleague work with him.

12 CONCLUSION

The points made in the abstract have been covered, though not as precisely as might be stated in a paper of a technical and mathematical nature, so reviewing those points here:

Very little was found in the literature about "people-maintenance".

The case examples showed dire results can follow poor people-maintenance, and from them we may infer that it's at least desirable, and, probably necessary, for the good of individuals and the employing organisation.

The difficulties in applying people-maintenance can be read from Section 8 above, where the individuals range from factory workforce to managers. Correction would have required change at senior management level.

The results of the seven case examples show the consequences of poor people-maintenance, and working back from those consequences the risk of losing employees and related costs can be seen. The other risk factor, the probability of each occurrence, is more difficult to assess, but in each organisation there was a sufficient build-up of factors to make the observed/reported outcome highly probable, if not certain.

The idea of condition-monitoring this form of maintenance is intriguingly attractive, and might require assessing some form of “happiness factor” of the employees, such as has recently been performed by Monash University through the Australian electorates. At this stage no further suggestions can be offered, it’s a topic for more detailed research.

Each of the case-examples showed costs from poor people-maintenance, some of that due to the crisis nature of what happened. If action had occurred to prevent the observed/reported outcomes there would have been some costs (for example, affecting cultural change isn’t cheap) but spread over a period, and the crisis impact would have been avoided.

Finally, the two examples in which a manager tried to apply good people-maintenance showed results which benefited both the individuals and the organisations involved.

Although the number of bad results exceeds the good outcomes in the above examples we conclude that good people-maintenance is desirable and should be applied where possible. Perhaps all that’s needed is for management to consider “what is right” (long term results) rather than “who is right” (which, in management’s mind, is usually the management).

13 REFERENCES

[1] Craggs WT. (c1945) Managerial Techniques. Sydney, NSW. CAC Apprentices Training School.

[2] Samson D. (1995) Management for Engineers. Melbourne, Victoria. Longman Australia Pty. Ltd.

[3] Herzberg F. (1966) Work and the Nature of Man. New York. World Publishing.

[4] Reason J. (1990) Human Error. Cambridge, England. Cambridge University Press.

[5] Mukhi S. Hampton D. and Barnwell N. (1991) Australian Management. Sydney, NSW. McGraw-Hill Book Company Australia Pty. Ltd.

EXPERIMENTAL STUDY ON MONITORING THE ATTACHMENT BOLT LOOSENESS IN A CLAMPING SUPPORT STRUCTURE MODEL

Dong G.M. [a,b] , Chen J. [a] and Zhang N. [b]

[a] *State Key Laboratory of Vibration, Shock& Noise, Shanghai Jiao Tong University, Shanghai, P.R. China.*

[b] *Mechatronics and Intelligent Systems Group, University of Technology, Sydney, Australia.*

Abstract: Based on the broad-band random vibration experiment on a structure model of clamping support used in transportation, this paper investigates the problems on detecting the time, evaluating the extent and locating the place of the attachment bolt looseness in the structure model. Firstly, wavelet transform was used to analyse the structural acceleration response data, which includes the moment when the bolt looseness occurred. An obvious step leap along the time axis, which shows the time when the bolt looseness occurs, can be clearly observed in the level-1 details of the wavelet decomposition. Secondly, ARX (auto-regression with exogenous inputs) model was constructed from the measured input and output signals, then the statistical indices of the model residual error were defined as the damage characteristic parameter indicating the severity and location of the loose bolts in the clamping support. The experimental study shows that the extracted characteristic parameter in this paper can be more appropriate than measuring the base frequency for monitoring the bolts looseness at early stage.

Key Words: broad-band random vibration experiment, clamping support structure model, attachment bolts, looseness, feature extraction, wavelet transform, ARX model

1 INTRODUCTION

Vibration is one of the main environmental factors which cause the product's malfunction. With the rapid technology development of the powerful power unit and the high-speed transportation utility, the vibration problem is more of concern. To guarantee the structure safety, especially the structure essential load bearing spot, or the vulnerable region which is unable to be directly inspected, the establishment of the damage monitor (examination), the early warning and the right moment service system is helpful to certain degrees in eliminating the hidden danger and avoiding the disastrous accident occurrence [1].

The commonly used local damage examination techniques (non-destructive evaluation (NDE)) have already been applied to the inspection of the structure crack, looseness and so on [2]. But for the structure real-time damage examination, the previous NDE techniques are not adequate. First regarding the large-scale component, it is very difficult to carry on the examination of all parts of the structure. Second, these technologies applied often request certain functions of the structure to stop using, which is very uneconomical. More serious is, once some important spots of the structure are damaged, the damage degree can develop very quickly, which can cause the entire structure to fail if the problem is not detected promptly. Therefore, it is necessary to develop the global based non-destructive evaluation technology [3], which is simple, economy, and does not affect the structure's normal operation.

The globally based non-destructive evaluation technology can be generally divided into two types: dynamics model based damage evaluation methods and signal analysing non-model based damage evaluation methods [4]. At present, there still is controversy in the dynamics model based methods to the practical applications, mainly because the validity of these methods relies on the accuracy of the analytical or the finite element model; but the dynamic performances of the actual complex component are influenced by the non-structural unit, environmental factor as well as boundary conditions; moreover the structure local variation is generally sensitive to the influence of the high order modes, which are difficult to measure accurately. Therefore, it is an important way to use the non-model based method for the structure overall situation monitoring by analysing and processing the structure dynamic response signal. In recent years, the good results are obtained by use of modern signal processing methods such as time series analysis [5], the wavelet transformation [6], and the high order statistical analysis [7] and so on.

To establish the damage monitoring system, the early warning and the on-line service system, there are two core technical issues: one is how to distinguish the occurrence of the structure damage, the extent of damage, and the damage location; the other is the life prognosis of the damaged structure system. This paper discusses the first issue based on the broad-band random vibration experiment on a structure model of clamping support used in transportation. Firstly, wavelet transform was used to analyse the structural acceleration response data, which includes the moment when the bolts looseness occurred. Secondly, ARX (auto-regression with exogenous inputs) model was constructed from the measured input and output signals, then the statistical indices of the model residual error were defined as the damage-sensitive feature indicating the severity and location of the loosening bolts in the clamping support. The experimental analysis shows that the extracted features in this paper are more appropriate than measuring base frequency for monitoring the bolts looseness at the early stage.

2 BROAD-BAND RANDOM VIBRATION EXPERIMENT ON THE STRUCTURE MODEL

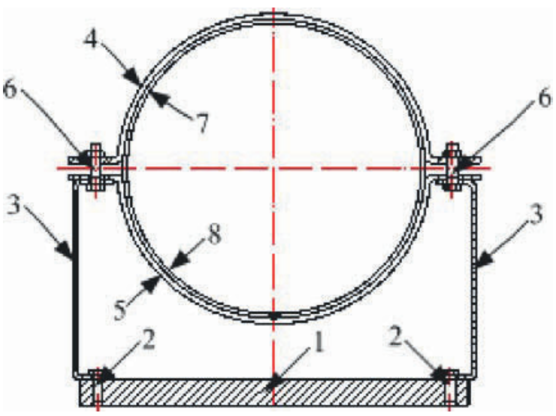


Fig1 Cross section diagram of clamping support

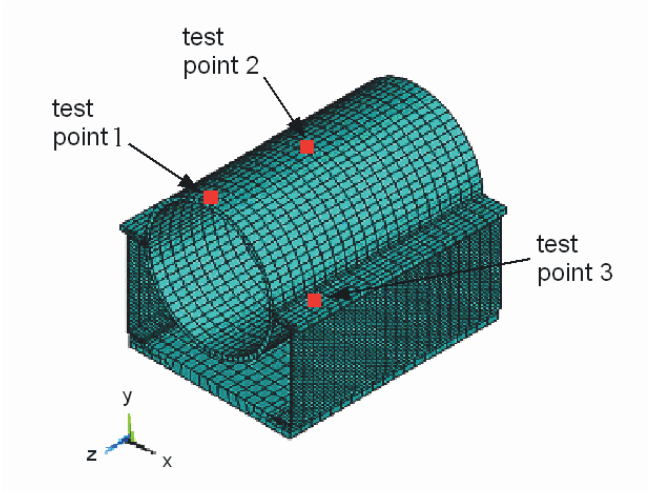


Fig 2 3D sculpt of clamping support and distribution of test points

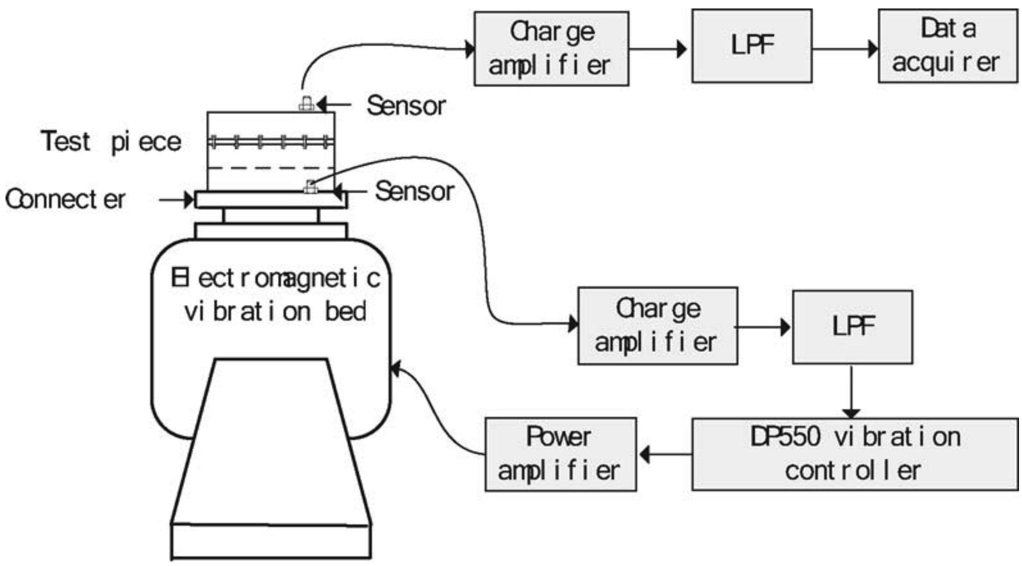


Fig 3 Sketch map of random vibration experiment system

The structure model of the clamping support is composed of a base, 2 blot-braces, a wrapper, 2 half steel-tubes and 24 bolts. The cross section diagram is shown in Fig. 1. On the bottom, the two rectangle blot-braces (No.3) are fixed on the base (No.1) by the 12 bolts (No.2), each side 6; On the top, the blot-braces are fixed with cylindrical wrapper (No.4, No.5) by 12 bolts (No.6), also each side 6. Due to bending pressure, the wrapper restricts the two half steel-tubes (No.7, No.8) inside. The general size of clamping support is: length (Z) 350mm, width (X) 280mm, height (Y) 258.5mm, thickness of blot-brace wall 3mm, thickness of wrapper 4mm, thickness of half steel-tube 3mm. In application, the bolts (No.2) connected with the base are always tightened. So in experiment we tighten and loosen the 12 bolts (No.6), which connected wrapper with blot-brace, to study the different dynamic response of clamping support under vertical broad band random excitation. The 3D sculpts of clamping support and the distribution of test points, at which the responses are taken, are shown in Fig. 2.

Many real vibration environments are stochastic: the uneven road surface causes the road vehicle vibration; the boundary layer turbulent flow causes the flight vehicle vibration. The random vibration experiment system includes two parts: vibration controlling subsystem and data acquisition subsystem. The former controls the vibration of electromagnetic vibration bed and is composed of sensors, charge amplifier, low pass filter (LPF), DP550 vibration controller, and power amplifier. The latter acquires data for analysing from test piece, which is connected with vibration bed platform through a connector. The data acquisition subsystem is composed of sensors, charge amplifiers, low pass filter and data acquirer. The sketch map of random vibration experiment system is shown in Fig. 3. There are four sensors in the experiment, three are placed at test point 1, 2, 3 shown in Fig. 2 and another sensor shown in Fig. 3 is placed on the connector for measuring input signal of the vibration bed.

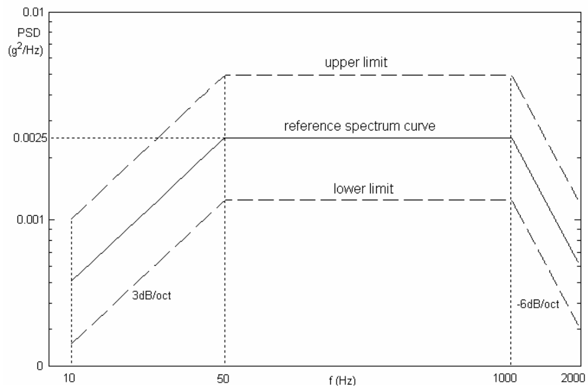


Fig 4 Reference PSD graph

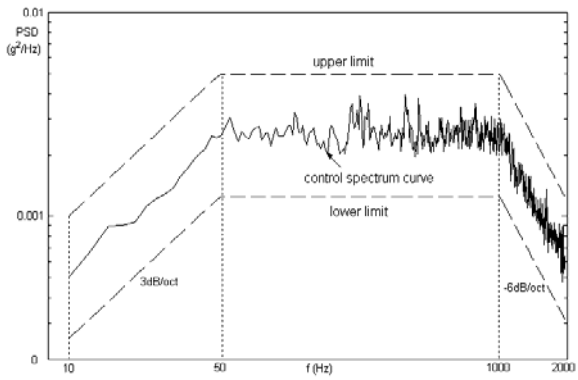


Fig 5 Actual control PSD graph

The random vibration experiment condition generally requires the power spectral density versus frequency curve (the PSD graph). The solid line in Fig. 4 shows the PSD reference spectrum curve which requests the vibration bed produce a straight spectrum in the 50-1000Hz frequency range. The solid line in Fig. 5 shows the actual PSD control spectrum curve. The dashed line in Fig. 4 and 5 shows the spectrum tolerance for upper and lower limits which are set as 3dB and -3dB relative to the amplitude of the reference spectrum. During the experiment, the vibration bed is run in the pre-experimental magnitude -6dB and -3dB for 20 seconds, then 20 seconds in the experimental magnitude 0dB, the sampling frequency is 7680Hz.

In order to simulate bolts looseness of clamping support with different degree, seven kinds of damage cases plus normal state (12 bolts (No.6) which connect wrapper and blot-braces are all tightened) are tested in the experiment. They are: (case1) 2 bolts closest to test point 3 are loosening completely; (case2) 2 bolts diagonal to test point 3 are loosening completely; (case3) 4 bolts at the four corners are loosening completely; (case4) 4 bolts at corners are tightened, the left 8 bolts are loosening completely; (case5) all the 12 bolts are loosening with slight degree; (case6) all the 12 bolts are loosening with medium degree; (case7) all the 12 bolts are loosening with serious degree. The object of this paper is to detect the time, evaluate the extent and locate the place of the attachment bolts looseness in the structure model under vertical broad band random excitation.

3 TIME DETECTION OF BOLTS LOOSENESS OCCURRENCE

Singularities and irregular sudden changes often carry the most important information in signals. In the mechanical structure damage monitoring, it is important to study the irregular components of the structure dynamic response, which imply the breakdown information of the structure. The Fourier transformation is one of the main mathematical tools for analysing singularities, but it provides a description of the overall regularity of signals and can not be well adapted finding the singularity locations of signals along time axis. The wavelet transformation can be seen as a windowing technique with variable-sized regions, which allow the use of longer time intervals for low frequency information and shorter intervals for high frequency information. One major advantage of the wavelet transformation is the ability to perform local analysis, through which the exact time of signal discontinuity can be detected [8].

There are literatures applying wavelet transformation to carry on irregularity detection of structure response signal, thus achieving the goal of structure damage monitoring. Hera [9], by simulation, applied the wavelet-based approach to study acceleration response data of a four-story prototype building structure subjected to simulated stochastic wind loading. He pointed out that the time of structural damage due to suddenly breakage of structural elements can be easily detected by spikes in the wavelet details. The damage region can be determined by the spatial distribution pattern of the observed spikes. In the authors' view, if a system is subject to broadband random excitation, the singularities could also be caused by the excitation process. Therefore, when the structure undergoes not the sudden damage but the cumulative damage in the actual stochastic vibration environment, we often cannot obtain the correct results. Masuda [10] studied the isolation method to separate the singularities due to the excitation and the singularities due to the system's changes through a simple single degree of freedom spring oscillator model, which has limited use in practical application.

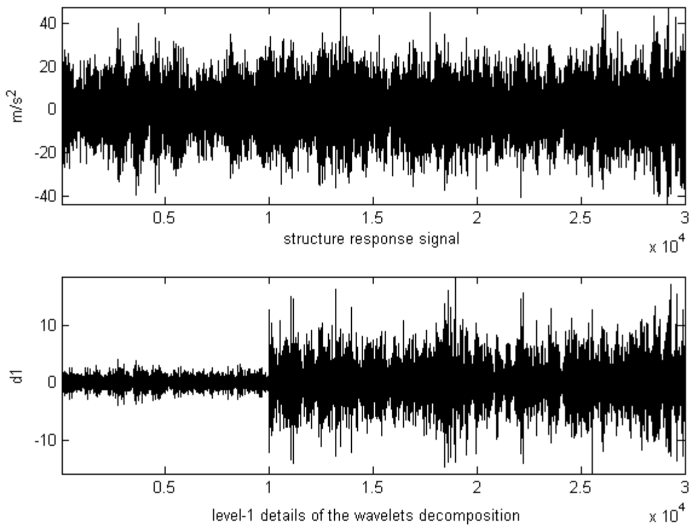


Fig6 Wavelet analysis

For detecting the time of the bolts looseness occurrence, this paper assumes that the bolts looseness is a sudden damage process. Using the Daubechies order 4 wavelets, the acceleration responses of the clamping support was analysed, which contained the occurrence of damage case 7. It is seen from Fig. 6 that, although there is no special change in the acceleration response signal, the obvious step leap can be seen in the first wavelet detail coefficient curve, which demonstrates the time of bolts looseness occurrence. Moreover, when analysing the data segment which did not contain the bolts looseness, regardless of the signal from the normal structure, or from the damaged structure, there is no obvious step leap in the wavelet detail coefficient curve, which is consistent with literature [9].

4 RECOGNITION OF BOLTS LOOSENESS DEGREE AND POSITION

The natural frequency of a structure system is one modal parameter that can be easily measured with high precision in the on-line monitor, therefore monitoring base frequency change is a simple and direct indicator of structure damage occurrence and degree [4]. However, this is not effective for the early small damage, which causes the necessity to seek more sensitive characteristic parameter to damage.

In the broad-band random vibration experiment, because the input energy of the vibration bed is very small, the power spectral density value 0.0025 g^2/Hz in the 50-1000Hz scope, the dynamic characteristics of the clamping support with attachment bolts looseness can be described using the linear model [11]. In this paper, the linear difference equation or the so called auto-regression with exogenous inputs (ARX) model [12] was used to express the system input-output relations.

4.1 Derivation of damage characteristic parameter

Firstly, divide the measured structure input and output signal evenly into M section, suppose the length of each data section is N. In the experimental data analysis of this paper, the actual signal is taken from 2 second to 18 second of experimental magnitude 0dB, and there are 29 data sections with length of each section 4096 points. Then the data normalization process is carried out for each section:

$$x_j(n) = \frac{\hat{x}_j(n) - \mu_{\hat{x}}}{\sigma_{\hat{x}}}, \quad n = 1,2,\ldots,N, \quad j = 1,2,\ldots,M \tag{1}$$

Where $x_j(n)$ is the normalized data section of $\hat{x}_j(n)$, $\mu_{\hat{x}}$ and $\sigma_{\hat{x}}$ are mean and standard difference value of $\hat{x}_j(n)$. The purpose of performing the data normalization is to reduce the influence of the external environment and the working condition changes on the extraction of system damage sensitive parameter.

Secondly, let $u_u(n)$ and $y_u(n)$ denote the input and output data section from the normal state structure after the data normalization, then establish the ARX model:

$$y_u(n) = \sum_{i=1}^{na} \alpha(i) y_u(n-i) + \sum_{j=1}^{nb} \beta(j) u_u(n-j) + e_u(n) \tag{2}$$

Where $\alpha(i)$ and $\beta(j)$ are model parameters, na and nb are model orders. To determine na and nb, the following steps are taken: first give an upper limit of the model orders according to the system physical property and the empirical knowledge, then calculate each candidate orders using the least square parameter estimation method, at last utilize the order decision criterion to determine the actual model orders na and nb. After computation, $ARX(60,60)$ model is used to express the input and the output relations of the normal state structure, and then the residual error $e_u(n)$ is a white noise sequence.

Thirdly, the same principle as above, obtain the input and output signal corresponding to the identical position of the structure with unknown condition (we don't know whether damage occurs). Let $u_d(n)$ and $y_d(n)$ denote the normalized input and output data section from above signal, then use the model parameters $\alpha(i)$ and $\beta(j)$ obtained from equation (2) to express the relations of $u_d(n)$ and $y_d(n)$, and the new residual error sequence $e_d(n)$ can be obtained:

$$e_d(n) = y_d(n) - \sum_{i=1}^{na} \alpha(i) y_d(n-i) - \sum_{j=1}^{nb} \beta(j) u_d(n-j) \tag{3}$$

Because the residual error represents the part that the ARX model cannot construct in the system output, we may believe that, if $u_d(n)$ and $y_d(n)$ comes from the same system with $u_u(n)$ and $y_u(n)$, then $e_d(n)$ can only have a smaller residual error; Otherwise, $e_d(n)$ will have a larger residual error.

Lastly, structure damage characteristic parameter can be defined as:

$$D_F = \sigma(e_d) / \mathrm{mean}(\sigma(e_u)) \tag{4}$$

Where $\sigma(e_u)$ and $\sigma(e_d)$ are standard differences of the residual error sequence $e_u(n)$ and $e_d(n)$, $\mathrm{mean}(\sigma(e_u))$ represents the mean value of $\sigma(e_u)$ corresponding to M sections from the measured input and output signal of structure in normal state.

Therefore, take the input-output signal from the structure of known normal state as the reference, through the damage characteristic parameter computation for each data section from the structure of unknown condition, damage characteristic parameter sequence $\{D_F(j),\ j = 1,2,\ldots,M\}$ can be obtained. Through the statistical analysis to the sequence, influence of non-linear factors such as the time varying characteristics of input can be decreased and robustness of damage diagnosis can be enhanced.

4.2 Recognition of bolts looseness degree

In this paper, the response signals of the clamping support measured from the test point 1 are calculated for different bolts loosening cases, which are (case1) 2 bolts loosening; (case3) 4 bolts loosening; (case4) 8 bolts loosening; and (case5, 6, 7) all the 12 bolts loosening with different degrees. The following table shows the computation results and the measured base frequencies of the clamping support with different damage cases:

Table 1: Mean and standard difference value of $\{D_F\}$ and base frequency

	case 1	case 3	case 4	case 5	case 6	case 7
Mean of $\{D_F\}$	2.8018	3.5678	4.8453	13.7342	19.4744	43.6092
Standard difference of $\{D_F\}$	0.3587	0.6612	1.2133	2.5648	3.4694	8.44
Base frequency (Hz)	217.5	213.8	206.3	185.3	167.8	132.7

Note: Measured base frequency of the clamping with normal state is 217.5 Hz

It can be clearly seen from above table that: as the bolts looseness degree or damage degree increases, the mean value of sequence $\{D_F\}$ presents an increasing tendency, therefore the mean value of $\{D_F\}$ can be taken as the characteristic parameter indicating the extent of damage. The standard difference of $\{D_F\}$ can be seen as an indicator of the credible degree of the sequence data and as a basis for comparing two groups of sequences.

It can also be seen from above table that: for the damage case 1, case 3, case 4, the corresponding base frequency change does not exceed 5%, especially for the damage case 1, when there are only two bolts loosening, the base frequency of the structure is same with that of normal state.

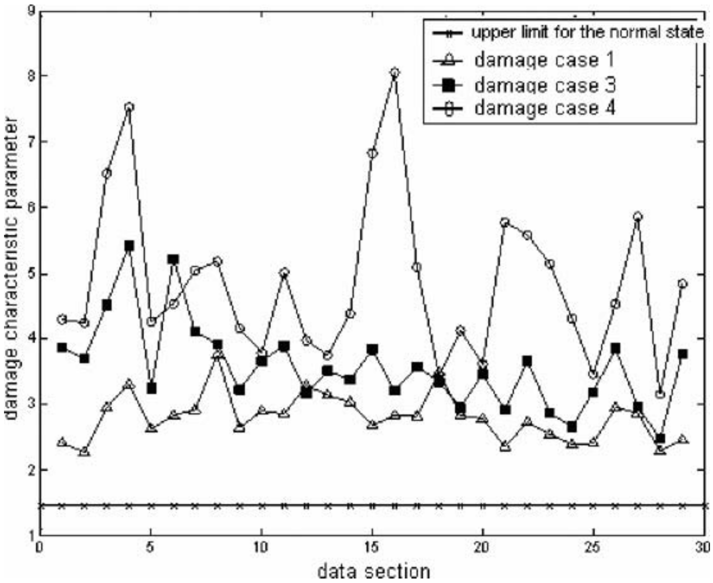


Fig 7detail of the $\{D_F\}$ sequences corresponding to the damage case 1, 3 and 4

The detail of the $\{D_F\}$ sequences corresponding to the damage case 1, 3 and 4 is shown in Fig. 7, from which we can see that even the sequence value for case 1 are obviously higher than the upper limit for the normal state. Supposes the sequence to obey the normal distribution, then we can calculate that the probability of $\{D_F\}_{case3} > \{D_F\}_{case1}$ is 84.57%, and the probability of $\{D_F\}_{case4} > \{D_F\}_{case3}$ is 82.24%, therefore three kinds of various degrees of bolts looseness conditions can be separated.

The above results are obtained from test point 1, and similar results are obtained from test point 2.

4.3 Recognition of bolts looseness position

In order to carry on the work of bolts looseness position recognition, the response signals obtained from the test point 1 and 3 are calculated for damage case 1 and 2, which have 2 bolts loosening with different position. The results are shown in the following table:

Table 2: Mean and standard difference value of $\{D_F\}$ for case 1 and 2

	Test point 1		Test point 3	
	case 1	case 2	case 1	case 2
Mean of $\{D_F\}$	2.8018	3.0348	4.1856	1.9729
Standard difference of $\{D_F\}$	0.3587	0.5334	0.5238	0.4331

It can be seen from above table that: because test point 1 is arranged on the top of the clamping support, the calculated structure damage characteristic parameter is insensitive to the variation of loosening bolts positions; however, because test point 3 is arranged near the attachment bolts, when the 2 loosening bolts is that nearest to the test point 3 (damage case 1), the mean value of $\{D_F\}$ is obvious higher than that of damage case 2, in which the 2 loosening bolts is far away from the measuring point 3. Therefore, if more sensors are laid beside each attachment bolt, the loosening bolts position may by roughly determined by the relative value of damage characteristic parameters calculated from each measuring point.

5 CONCLUSIONS

(1) Compared with the Fourier transformation, the wavelet transformation has the ability to perform local analysis, through which the exact time of signal discontinuity can be detected. Experiment analysis shows that an obvious step leap along the time axis, which shows the time when the bolt looseness occurs, can be clearly observed in the level-1 details of the wavelet decomposition.

(2) ARX model is constructed from the measured input and output signals, and the statistical indices of the model residual error were defined as the damage characteristic parameter. Based on damage case 1, 3, 4, 5, 6 and 7, the damage characteristic parameter rightly indicates the bolts looseness degree of the clamping support structure model. Based on damage case 1, 3 and 4, the damage characteristic parameter is more sensitive than measuring base frequency change for monitoring the bolts looseness at the early stage.

(3) Based on the analysis of damage case 1 and 2, we can envisage that if more sensors are laid beside each attachment bolt, the loosening bolts position may be roughly determined by the relative value of damage characteristic parameters calculated from each measuring point.

6 REFERENCES

[1] J. Chen. (1999) Vibration monitoring and fault diagnosis of mechanical systems. Shanghai: the Shanghai Jiao Tong University Press. (in Chinese)

[2] C.Z. Chen, et al. (2001) Structure damage examination and intelligent diagnosis. Beijing: the Science Press. (in Chinese)

[3] D.W. Zhang & P.X. Wei. (2000) Model updating and damage identification. Beijing: the Science Press. (in Chinese)

[4] S.W. Doebling, C.R. Farrar & M.B. Prime. (1998) A summary review of vibration-based damage identification methods. *The Shock and Vibration Digest,* 30(2), 91–105.

[5] H. Sohn & C.R. Farrar. (2001) Damage diagnosis using time series analysis of vibration signals. *Smart Materials and Structures*, vol.10, 446-451.

[6] A.N. Robertson, C.R. Farrar & H. Sohn. (2003) Singularity detection for structural health monitoring using holder exponents. *Mechanical Systems and Signal Processing*, 17(6), 1163-1184.

[7] J. Chen, et.al. (2002) A bispectrum feature extraction enhanced structure damage detection approach. *JSME International Journal, Series C*, 45(1), 121-126.

[8] A. Al-khalidy. (2002) Health monitoring of dynamic systems using wavelet analysis. Ph.D. Dissertation, Department of Civil and Environmental Engineering, Cornell University.

[9] A. Hera & Z.K. Hou. (2004) Application of wavelet approach for ASCE structure health monitoring benchmark studies. *Journal of Engineering Mechanics*, 130(1), 96-104

[10] A. Masuda, et. al. (2002) Wavelet-based health monitoring of randomly excited structures. In Andrew Smyth(Ed) *15th ASCE Engineering Mechanics Conferences*, Columbia University, New York, NY, pp1-8.

[11] S.S. Ivanovic, M.D. Trifunac & M.I. Todorvska. (2000) Ambient vibration tests of structures - A review. *Indian Society of Earthquake Technology Journal*, 37(4), 165-197.

[12] L. Ljung. (2002) System Identification--Theory for the User, second edition. Upper Saddle River, N.J.: Prentice Hall.

Acknowledgments

The presented work was financially supported by National Natural Science Foundation of China and Institute of Engineering Physics of China (approved No. 10176014), and partially by University of Technology, Sydney.

APPLICATION OF PROBABILISTIC SAFETY ASSESSMENT TO PARAMETERS OF OPERATIONAL LIMITS AND CONDITIONS OF THE OPAL RESEARCH REACTOR

Bastin, S.J., and Perera, J.

Australian Nuclear Science and Technology Organisation, New Illawarra Road Lucas Heights, NSW 2234 Australia.

Abstract: This paper gives a brief explanation of Probabilistic Safety Assessment (PSA), the meaning and application of Operational Limits and Conditions (OL&Cs) for facilities and describes the application of PSA to the derivation of OL&C parameters such as Completion Times. ANSTO's Research Reactor OPAL is used as a case study but the authors conclude that the methodology, based on NUREG/CR-6141, could equally be applied to major hazard facilities.

The PSA indicates that the OPAL reactor meets the regulatory safety limits and safety objectives. The very low risk is achieved by important design characteristics such as the two independent and diverse shutdown systems, minimal dependence on support systems for the safety functions, the absence of in-core experiments, and the redundancy in cooling modes.

Completion times are derived for safety systems and some sub-systems by balancing the incremental risk of single outages of those systems, and also by balancing the incremental risks associated with the expected number of outages of those systems and subsystems while ensuring that the incremental risks are tolerable.

This methodology can give an estimate of the total incremental risk of allowable outages of safety systems and subsystems taking into account the expected number of outages of those systems each year. An assessment of the tolerability of that additional risk can be made.

Key Words: Probabilistic Safety Assessment, Operational Limits, OPAL, Dependence, Tolerability.

1 INTRODUCTION

Understanding and managing the risks of operations is a critical aspect of asset management and a legal requirement on owners and managers of plant. To assure and demonstrate the safety of complex or hazardous plant, it is important to undertake Probabilistic Safety Assessmenti (PSA) in addition to the Deterministic Safety Analysis (DSA). PSA can provide additional insights into the strengths and possible weaknesses of a plant design, and can give indicative estimates of the risk associated with the plant. The application of PSA was once confined to nuclear power reactors but is becoming more common for all nuclear facilities, and has also been applied, at least in some limited forms, to complex and major hazard plant, especially in petrochemicals, and oil and gas production [ii, iii, iv, v].

PSA is a valuable tool for the quantification of risks, the identification of possible plant responses to initiating events, and the identification and ranking of features that govern the risk.

Yet following commencement of operations of a plant, the risks are often altered from that modelled in the PSA. This can happen in the following ways:

a) When failures in some component of safety systems occur, some leeway is often sought by the operating organisation to correct the fault while still in operations.

b) Over time, some of the assumptions made in the development of the PSA can change. For example, the test frequency of some systems, subsystems or components may change for practical operational reasons.

c) More accurate data on equipment and human reliability may become available in the operational phase.

d) Plant modifications in the operational phase may require changes to the PSA models.

The approach to resolving these issues in the nuclear industry has been (i) to use the PSA to assist in the derivation of acceptable amount of "leeway", (ii) to use the PSA to help choose appropriate test intervals and (iii) to keep the PSA as a "living" document that is used to assess the altered risk of proposed changes, and is then updated to reflect the changes once implemented.

This paper discusses some of the techniques used, using the OPALvi research reactor as a case study, and proposes that there may be applications for non-nuclear plant in the growing field of Engineering Asset Management.

2 OBJECTIVES OF THE OPAL PSA

The OPAL reactor is an open pool type, that is, a reactor where the core sits in a deep pool of water that provides cooling of the core, and shields people and equipment from the harmful radiation. The metallic pool liner sits in a high integrity reinforced concrete block. The OPAL reactor will provide facilities for irradiating targets for the production of radiopharmaceuticals, and silicon and for experiments, as well as providing high quality neutron beams for specialised research.

The reactor vendors, INVAP SE, provided a PSA for licensing purposes for the OPAL research reactor [vii]. This PSA underwent significant review by ANSTO and supported the application for an operating licence to the Australian Radiation Protection and Nuclear Safety Agency (ARPANSA).

In addition to the general objectives of PSA (mentioned above), the main objective of the OPAL reactor PSA is to support the deterministic safety analysis in demonstrating the safety of the plant and demonstrating compliance with probabilistic safety limits and objectives. Moreover, since this PSA was developed in parallel with the basic engineering phase of the OPAL reactor, the preliminary results were used as input to the design process, thus permitting improvements to be made to the design.

The OPAL PSA is considered Level I, with some Level III considerations. Level I PSA examines the risks of breaching the primary barrier for radioactive material *eg* melting fuel, or damaging the fuel cladding. Level II PSA examines the risk of release of radioactive material to the environment, while Level III PSA examines the risk of radiological consequence to people. Probabilistic objectives and limits for doses to members of the public are set by ARPANSA.

In non-nuclear applications, analogous primary measures of risk need to be identified, and in many cases, the three level concept may be applied although the distinction between Level II and Level III may not always be warranted. For example, in a facility which processes and stores hazardous chemicals, the Level I measure may be the "breach frequency" of the storage and process vessels and pipework. The Level II PSA may analyse the possible chemical and physical interactions in a breach, leading to the frequency of releases of different magnitudes to the environment. The Level III measures will be the frequency and severity of exposures of those released chemicals to employees, contractors and members of the public.

3 ANALYSIS METHOD

This section describes the PSA methodology using the OPAL PSA as a case study.

3.1 PSA Approach

PSA involves combining event tree analysis and fault tree analysis to model the propagation of accident sequences beginning with Initiating Events (IEs) and subsequent failures of defence systems. The likelihood of different groups of outcomes (referred to as end states), are then quantitatively analysed using probabilistic calculations. Event tree analysis forms the backbone of this process. Each IE has a corresponding event tree in which the applicable defence systems, are modelled as the nodes of the event tree. Fault trees are used to model the IEs and also the probability of failure of each of the defence systems. The initiating events and the end states are both calculated as frequencies (usually per year).

3.2 Identification and Selection of Initiating Events

The identification and selection of IEs always poses a question as to the completeness of the PSA, and no method can guarantee completeness, although use of one or more standard methods gives reasonable assurance. A systematic method, known as Source and Event Analysis (SEA[viii]) was used for the OPAL PSA. This method is consistent with NUREG/CR-2300 [ix] and the approach given in IAEA-TECDOC-517 [x]. The SEA method is a multi-step, bottom-up approach that involves identification of the sources of radioactive material, the barriers that separate these sources from the public and/or plant personnel and the primary failure mechanisms of these barriers. From these mechanisms, IEs can be postulated. Those events which make a negligible contribution to the overall risk, for example, because the inventory of radioactive material is too small to pose a risk are screened out according to internationally recognised guidelines [xi].

3.3 PSA Models and Data

Many sophisticated tools for the quantification of risks arising from the operation of nuclear reactors and other complex installations are now available. SAPHIRE [xii] is a package used to quantify Level I PSA (*ie* to estimate the frequency of the major undesired event – core damage) and was used for the OPAL study. Industry-specific tools and software must be used to assist in the modelling of dispersion and consequence. For the OPAL study, such modelling was carried out using PC Cosyma (developed by the UK National Radiological Protection Board under a European Commission contract).

Once the IEs are identified, the following steps are typically undertaken in PSA:

a) Each IE is analysed and the possible accident sequences derived from it are mapped against the possible defences (mainly safety systems) that would be required along those sequences.

b) A System Fault Tree (FT) is developed for each of the required defences, using the system success criterion, and relevant drawings and the Operational Limits and Conditions.

c) An Event Tree (ET) is developed for each IE, with the corresponding nodes of the tree corresponding to a defence system. Sequences which have no physical meaning are removed.

d) The ETs and FTs are programmed into a PSA software package. In the OPAL PSA, the SAPHIRE [xii] package was used. The PSA model is then checked and the qualitative results examined. The model is then quantified. In the OPAL PSA, equipment failure rate data used was mostly from the IAEA databases [xiii,xiv].

e) The frequencies of the main plant damage states are determined by the sum of the frequencies of sequences that lead to those damage states. For the OPAL PSA, core damage is the main plant damage state of interest and the results focussed on that, although other release sequences were identified.

f) Importance measures are estimated to give an indication of how significant each basic event is to the overall core damage frequency (CDF). Importance measures are an indication of sensitivity. In the OPAL PSA, it was also of interest to look for "cliff-edge" effects – that is, situations where a small change in basic event data can cause a significant increase in the overall CDF.

g) The probability of each basic event has some associated uncertainty which is represented by the associated confidence interval on the data. Having obtained "best estimates" of all of the end state frequencies, it is usually of interest to understand the uncertainty associated with these estimates. This is achieved by "propagating" the uncertainties through the Level I PSA model. The objective is twofold: (1) to understand the limitations of the numerical results and (2) to obtain upper bound estimates of the end state frequencies at specific levels of confidence.

h) In a full level III PSA, the response of containment and the dispersion of associated releases are modelled probabilistically. In the OPAL PSA, such detailed modelling was not warranted. However, to estimate the plant behaviour beyond the scope of the Level I PSA, a set of representative accidents were analysed to determine inventory, containment response, and consequences of such a release. These accident sequences which included those not related to core damage, were those estimated to have annual frequency close to or within the range of ARPANSA risk acceptance criteria for radiological consequence, and bounded other credible accident scenarios of lesser significance. Using conservative retention factors, and conservative assumptions about the containment response, source terms were derived for each of the grouped plant damage states and these source terms were used to estimate the doses a member of the public might receive, making conservative assumptions about the weather conditions, and sheltering etc. The results were then compared against acceptance criteria.

3.4 Dependent Failure Analyses

Actual operating experience with complex plants demonstrates that, although the likelihood of a series of failures is quite small, this likelihood is numerically higher than would be estimated solely from a postulated chain of independent failures. This is because physical and human interactions result in dependent failures that increase the conditional probability of each successive failure in the chain. The treatment of dependencies in the identification and quantification of accident sequences is called "dependent-failure analysis". Since essentially all important accident sequences that can be postulated for nuclear reactor systems involve the postulated failure of multiple components, systems, and containment barriers, dependent-failure analysis is an extremely important aspect of PSA. This will no doubt be true for non-nuclear major hazard plant.

The main focus of this paper is not on dependent failures analysis but it may be warranted to briefly explain how PSA in the nuclear industry traditionally deals with dependencies. A detailed engineering analysis of dependent failures must consider the root causes of component failures and the degree of dependence among component failures with regard to each root cause. This is particularly important for a new reactor, where no operating, maintenance and testing experience exists. There are three types of dependent failures:

a) <u>Functional dependencies between systems</u> are where the success of one system is dependent on the success of another, for example due to reliance on a support system or where there is a shared component or subsystem in two systems. Functional dependencies are treated explicitly in event trees while shared component dependencies are taken account of by the analysis software, providing the shared items are clearly identified by the analyst.

b) Dependencies or common causes between basic events can sometimes be foreseen (*eg* those caused by improper maintenance or incorrect restoration after testing) and these can be modelled specifically where identified. However, it has been recognised that many significant dependencies cannot be foreseen. Traditionally, these common cause failures (CCF) are modelled as a numeric fraction of the basic event failure probability (*eg* the beta factor, or multiple Greek letter methods). Parameters for these models make use of generic data, which are then updated with plant-specific data, where available. For the OPAL PSA, in lieu of the more conventional method, common cause effects were modelled by recognising that ultimately all faults can be attributed to human error [xv]. Because this method is novel, comparisons were made with conventional beta factor and MGL methods and the effective CCF contributions were similar to those that would have arisen from conventional methods.

c) Dynamic human interactions cover the errors that operators could make in responding to an IE. In the OPAL reactor, the actuation of the safety systems is automatic, and no credit was taken for manual actions or recovery actions. It was difficult to anticipate how an operator might respond erroneously in a particular sequence, when no operator response is required in the first 30 minutes, however, where credible operator actions that could jeopardise safety system functions were identified, they were included as basic events in the fault tree models.

3.5 Seismic modelling

Again, seismic modelling is not the main focus of this paper, but it may be useful to understand how PSA in the nuclear industry deals with seismic issues. There are two aspects to seismic analysis in PSA: (1) obtaining measures of the seismic hazard at the site, usually characterised by the zero period peak horizontal ground acceleration (PHGA) at various return periods, and (2) obtaining measures of the fragility of systems and components of the plant which are required in responding to seismically induced IEs [xvi]. The seismic hazard is considered the initiator of several possible accident sequences, such as a loss of coolant accident (LOCA) or loss of flow. The ability of the safety systems to mitigate the consequences of the initiated fault condition depends on the extent to which seismically induced failures of components have degraded the functional capability of those systems. Failure probabilities (conditional on a specific level of ground acceleration) of components are derived from their seismic fragilities.

The seismic hazard for Lucas Heights was estimated by consultants [xvii] in 1999 and refined by a broader group of experts in 2001 [xviii]. Seismic PSA assumes that the designers have adequately taken into account the spectrum of the seismic ground accelerations so that the seismic PSA can focus on one main parameter of the hazard (*ie* the zero period PHGA) at each return period. Fragilities were derived using commonly adopted values for the uncertainty ($\beta_R = 0.25$; $\beta_U = 0.35$) and setting the high confidence low probability of failure (HCLPF) as the mean 10,000 year ground acceleration for seismic category 1 items and systems. This is appropriate, because all seismic category 1 items and systems are designed with high confidence to continue to function under such ground acceleration.

The seismic contribution to the CDF has been analysed in two different scenarios. The first scenario considers the contribution to the CDF for those seismic events whose frequency is up to the 10,000 year earthquake. However, because all safety critical items have been designed to this level of earthquake, the estimated additional CDF from seismic events up to this level is negligible. For this reactor, the CDF contribution from seismic events up to the 10,000 year earthquake is insignificant and so is not discussed further.

The second scenario considers the contribution to the CDF for all the seismic events indicated by the full Hazard Curve for the site (*ie* up to 100,000 year return period).

4 LEVEL I PSA RESULTS

4.1 Numerical results

4.1.1 Core damage frequency

Significant core damage (such as melting or rupturing of the fuel cladding) can lead to a release of fission products to the coolant and ultimately to the containment building. Degradation of containment can release fission products to the environment. In the context of Level I PSA of a nuclear reactor, the prime measure of risk is the CDF, that is, the expected frequency of significant damage to the reactor core. This is obtained by the summation of all the event sequence frequencies that lead to core damage. The CDF is usually divided into the CDF from sequences arising out of internal random failures, internal fire and flooding, external events excluding seismic events, and seismic events.

Table 1 shows the overall CDF with 90% confidence limits, including all identified external events with the full seismic hazard[vii]. It can be seen that the estimated CDF is remarkably low demonstrating the robust design of the OPAL reactor.

Table 1 Overall Core Damage Frequency including internal, external and using the full seismic hazard from reference [vii]	
Mean Core Damage Frequency:	$5.4x10^{-7}$/year
5^{th} Percentile Core Damage Frequency	$3.0x10^{-7}$/year
95th Percentile Core Damage Frequency	$9.1x10^{-7}$/year

The relative contribution to the mean total CDF due to each of the main classes of initiating events is indicated in Figure 1.

It can be seen that the seismically induced loss of offsite power is the greatest contributor to the CDF with (about 48% of CDF) and seismically induced LOCAs the next most significant cause. It should be remembered that this is based on several conservative assumptions: (1) that seismic events above the one in 100 year event are guaranteed to cause loss of offsite power and (2) that no credit is given for operator intervention. In reality, with a LOCA in progress, operators would ensure the LOCA is arrested by clearing any blockage in the siphon effect breaker, forcefully opening a flap valve or shutting the manual isolation valves. Loss of offsite power (non-seismic) and internal (or randomly occurring) LOCAs are the next two most significant classes of initiators.

4.1.2 Comparison with regulatory requirements

The Regulatory Assessment Principles expect for a new nuclear reactor, the applicant to be able to demonstrate that the frequency of significant damage to the nuclear core is less than 10^{-5} per year. Furthermore, dose-based safety objectives specify the expected total frequencies in various bands of dose to members of the public. Of these dose-based safety objectives, the most stringent (*ie* the expected total frequency in the most severe dose category) is 10^{-6} per year [xix]. The values of CDF shown above demonstrate that the Reactor fulfils the most stringent Safety Limits and Objectives, with 95% confidence. Therefore, it can be stated that those accidents with the potential to cause a significant damage to the core, pose a negligible risk to the public in the vicinity of the reactor.

4.2 Design improvements

The OPAL reactor has two independent and diverse reactor protection systems and corresponding shutdown systems. Whilst both are essentially equal in their respective capability to respond to all design basis events, preference is given to the first shutdown system (FSS) because of the greater time required to reset the second shutdown system (SSS). The second reactor protection system (SRPS) was therefore designed with a smaller set of trip parameters than the first protection system (FRPS). In the preliminary stages of the design, the preliminary PSA results showed that loss of heat sink (*ie* loss of the secondary cooling system function) together with failure of the FRPS was a significantly dominant sequence. Whilst this sequence was still predicted to be extremely unlikely, the PSA showed that the design was not well balanced, in that the total internal CDF was dominated by just one sequence. On the basis of the PSA results, a design decision was taken to implement an additional SRPS trip to respond to the loss of heat sink initiator.

5 LEVEL III PSA RESULTS

It is usual that the accident scenarios that contribute the most to the risk of a nuclear reactor are those that involve substantial damage to the reactor core. However, the very robust design of the OPAL reactor, with inherent safety features and a high degree of redundancy, diversity and independence of its safety functions, makes the CDF so low, that they are not considered credible. One of the reasons for this is that the PSA was developed at the same time the basic engineering was developed, and any weak points identified during the system's analyses for the PSA, were passed to the designers to implement improvements [vii]. As a result, for the OPAL reactor the risk representative scenarios do not involve significant core damage. These other accident scenarios relate to target damage or fuel damage outside the core and they were analysed and quantified both in their frequencies and their associated potential doses on the public. The release categories (RC) for the representative release events, even when modelled under extremely conservative assumptions, show that the expected risk contribution of the selected accidents is well within acceptable levels.

It is important to notice also, that the maximum doses do not require any off-site emergency measure (eg sheltering).

6 APPLICATIONS OF PSA

The OPAL reactor PSA in its current form was primarily intended for demonstrating compliance with regulatory objectives. However, it has potential further practical and beneficial uses such as safety assessment of any future modifications of the reactor and risk-informed decision making on reactor operational aspects. An application of the OPAL reactor PSA is to derive risk-based parameters for the Operational Limits and Conditions (OL&Cs) is described below.

6.1 Completion times

When a failure occurs in a system important to safety, the risk is increased above the originally calculated risk (e.g. CDF) until that failed item is restored. By choosing an appropriate value for the time in which restoration is allowed, the increased risk can be kept within agreed acceptable bounds, both for the single event and for the total number of expected such events in one year.

A Completion Time is the allowed time specified in the OL&Cs to achieve a required action following a failure in, or detection of the inoperability of, a safety system. The required action is intended to put the reactor in a safe or safer state *eg* by restoring a failed item, setting a failed instrument to the "tripped" status, or simply shutting down the reactor . If the required action is fulfilled within the completion time, the OL&Cs are deemed to have been complied with. Traditionally, completion times are derived from experience and past practice but the use of PSA permits risk-informed derivation of completion times.

For the OPAL reactor, the NUREG/CR-6141 methodology [xx] was followed as far as was possible. Using the risk increase and risk reduction factors for systems and components, it was possible to determine the effective increased risk of remaining in a state with a system or component inoperable. By comparing that incremental risk with some criteria, a maximum completion time can be determined. This permits a balanced approach to determination of completion times. Knowing how often a system is likely to come into a degraded or inoperable state and comparing the increased annual risk permits derivation of a second measure of the allowable completion time. The actual completion time should not be greater than the minimum of these two measures. The following outlines the method in more detail.

1. The first step is to define criteria for the acceptable additional risk of core damage allowed for a single event ($\boldsymbol{r_c}$), for the additional annual risk associated with the expected number of failures or outages for each system ($\boldsymbol{R_c}$) and for the total expected number of failures in all systems ($\boldsymbol{R_{cT}}$).
2. Next, one must identify the systems, subsystems and components for which completion times are required.
3. For each of the systems, subsystems and components, obtain the Risk Increase Factor, ***RIF***, and Risk Reduction Factor, ***RRF*** from the PSA. The ***RIF*** and ***RRF*** are importance measures of events in the PSA and are typically computed by the PSA software. The ***RIF*** is the factor by which the CDF increases if the event concerned has definitely failed. The ***RRF*** is the factor by which the CDF reduces if the event concerned is definitely successful i.e. not failed. For coherent systems, both the ***RRF*** and ***RIF*** are greater than or equal to 1.
4. Using the baseline CDF, ***R(x)***, the Risk Increase Factor, **RIF**, and Risk Reduction Factor, **RRF**, for the relevant system or component, taking into account any adjustments for the relevant degradation rather than failure of the system, calculate the per unit time additional risk of an outage (**Δ*R***).

$$\Delta R = \left(RIF - \frac{1}{RRF} \right) R(x) \qquad \textbf{(1)}$$

5. Using failure rate data from the PSA or from other relevant data sources, estimate the summed failure rate of components in the system to estimate the outage frequency (***f***).
6. Compute d_1, d_2 and d_{min} as per equations (2), (3) and (4) below where min(x, y) is the minimum of x and y.

$$d_1 = \frac{r_c}{\Delta R} \qquad (2)$$

$$d_2 = \frac{R_c}{\Delta R f} \qquad (3)$$

$$d_{\min} = \min(d_1, d_2) \qquad (4)$$

7. Calculate the sum of the additional annual risk for each system $\boldsymbol{R_T} = \Sigma \boldsymbol{R_c}$ and if this exceeds $\boldsymbol{R_{cT}}$ then reapportion the risk and repeat.

6.2 Criteria

For the OPAL Reactor, criteria for the derivation of risk-informed completion times were determined by setting the allowed outage time for the first shutdown system to 0.5 hours. This approach, although somewhat arbitrary, was a way of obtaining a feasible completion time while still keeping the additional risk to a level about an order of magnitude lower than the full CDF.

The results were r_c = 7.94x10^{-8} (dimensionless) and R_c = 6.67x10^{-8} pa. The criterion r_c is the acceptable increased probability of a single event that continues for the full allowed completion time. The criterion R_c is the acceptable increased annual CDF frequency arising from the expected number of outage events assuming each continues for the full allowed completion time.

6.3 Systems, Subsystems and Components

Completion times were required generally for any system that helps prevent core damage (as described in the PSA) or helps mitigate consequences. For some systems, completion times were required for subsystems (rather than components). Risk Increase Factors (RIFs) and Risk Reduction Factors (RRFs) are generally only available for components – not for systems – in the current PSA (the reactor protection systems being the exceptions. However, most systems have a common cause event at the highest level (thus forming a single order cut set for that system). The RIF and RRF for such events should be precisely the RIF for the system. Similarly, the RRF for the CCF event should be a reasonable approximation for the RRF of the system.

For some systems, the completion time for a partially degraded system was required (for example, the reactor protection systems, with one channel of a particular function inoperable). A slightly different but similar approach was adopted but this is beyond the scope of this paper.

As discussed above, a slightly different approach was required for those systems that do not contribute to CDF (but have a mitigative effect if core damage were to occur). It was assumed that failure of these systems coincident with core damage would lead to releases ten times worse than with all such systems functional.

6.4 Results of the Completion Times Analyses

The results of the Completion Time derivation are shown in the table below.

Table 2 Completion times for systems and channels/trains

	Risk-Based Completion Times (hours)			
System	**by single event**	**by annual risk**	**Minimum**	**Proposed**
First Reactor Protection System	11.9	902	11.9	12
Second Reactor Protection System	2.16	288	2.16	2
First Shutdown System	0.50	0.50	0.50	0.5
Second Shutdown System	2.15	1.74	1.74	1.5
Flap Valves	5.49	1158	5.49	5
Emergency Makeup Water System	1376	2408	1376	1200
Standby Power System	1035	538	538	528
Containment Isolation	550	1080	550	528
Containment Energy Removal System	548	33	33	32
FRPS single channel	351	212	212	192
SRPS single channel	200	59.2	59.2	60
Single Flap Valve	9068	1912761	9068	1200
SPS Single Diesel	17925	15132	15132	1200
CI Single Train	8005	95532	8005	1200
CERS Single Train	6449	227	227	216

Note: The values in the "proposed" column above were derived by risk based methods with some rounding down (to nearest hour, day or week) and the reduction of very large values. As the risk criteria were set arbitrarily low, there is significant flexibility in the final values which were finalised taking into account other technical criteria and practical considerations. These values are indicative only and not to be taken as the final decision on completion times.

6.5 Test Intervals

Where generic probabilities of failure on demand were used for items in safety systems, maximum test intervals were derived by applying the unrevealed/test interval model to the probability of failure. The adoption of such test intervals would ensure that the safety systems will have the required reliability consistent with the data used for the PSA.

To calculate the test intervals, for single components and single channels the following formula was used:

$$p = \frac{\lambda\tau}{2} \qquad \textbf{(5)}$$

This makes use of the small failure rate / small test interval assumption. The equation is then solved for the test interval, τ.

$$\tau = \frac{2p}{\lambda} \qquad \textbf{(6)}$$

From the above data and equations, test intervals were found for all the relevant components.

For the First Reactor Protection System (FRPS) and Second Reactor Protection System (SRPS), because the PSA assumed a single probability of failure on demand for the system (or relevant system function), the test interval was calculated at a single channel level, assuming that a channel failure is dominated by failure of the sensor/transmitter.

$$p = 3 \cdot \left(\frac{\lambda\tau}{2}\right)^2 \text{, thus} \quad \tau = \frac{2}{\lambda} \cdot \sqrt{\frac{p}{3}} \qquad \textbf{(7)}$$

However, since any set of three instruments on a trip parameter would have all three instruments tested at the same time, the instantaneous unreliabilities of each channel are synchronised, and the average of the product of any two leads to a slightly different denominator as follows:

$$p = 3 \cdot \left(\frac{\lambda^2\tau^2}{3}\right) = (\lambda\tau)^2 \text{, thus} \quad \tau = \frac{\sqrt{p}}{\lambda} \qquad \textbf{(8)}$$

Note: For each of the FRPS and SRPS it was also conservatively assumed that only one input signal parameter (with 3 channels in 2oo3 voting configuration) is capable of the detection of a fault sequence, although in practice for many sequences more than one parameter will be capable of detecting and responding.

For the Diesel Generators, although there are three diesels, this system model was simplified to consist of two symmetrical trains. The system probability of failure on demand was derived in the PSA so the test interval was derived as follows assuming synchronised testing.

$$p = \left(\frac{\lambda^2\tau^2}{3}\right) \text{, thus} \quad \tau = \frac{\sqrt{3.p}}{\lambda} \qquad \textbf{(9)}$$

6.6 Results of Test Intervals Analyses

The results of the Test Interval Analyses are shown below.

Table 1 Test/inspection or surveillance intervals for systems and components

System	Component	Calculated Test Interval (h)	Proposed Test Interval (h)
First Reactor Protection System	Instruments	29900	13140
Second Reactor Protection System	Instruments	9550	8760

System	Component	Calculated Test Interval (h)	Proposed Test Interval (h)
First Shutdown System	Valve	533	840
	Rod	400	840
	Compressor	12000	168
Second Shutdown System	Ball valve	65100	2400
	Solenoid valve	2400	2400
Flap Valves/Siphon Breakers		870	840
Emergency Makeup Water System	Float valve	870	840
	Ball valve	4000	840
Containment Energy Removal System	Fan A	189	168
	Compressor	2200	
	Fan B	5660	168
	Pump	1330	
	Valve	294	
Containment Isolation System	Valve	12600	2400
	Solenoid valve	2410	2400
	Check valve	980	840
Standby Power System	Diesel generator	3030	840

Note: The values in the proposed column above were derived by risk based methods with some rounding down (to nearest hour, day or week) and the reduction of very large values. These were to be considered along with other technical criteria in the finalisation of the OL&C surveillance times. These values are indicative only and not to be taken as the final decision on test intervals.

7 CONCLUSIONS

A Level I PSA with some Level III considerations was conducted on the OPAL Reactor and shows that the reactor is safe and meets regulatory objectives. The PSA has been used to determine risk-based parameters for Operational Limits and Conditions and will continue as a living PSA for the assessment of modifications and other applications. This application of PSA is now standard for nuclear power plant and is becoming more common in other nuclear facilities such as research reactors. Furthermore, this application of PSA is of significant practical benefit in guiding decision makers on the appropriate values for parameters such as completion times and test intervals and can of benefit in demonstrating to regulatory bodies that these values are appropriate. Since probabilistic safety assessment, in some form or another, is becoming more common in non-nuclear facilities (such as petrochemicals and oil and gas production), it follows that there may be similar applications to operational limits and conditions in such facilities.

8 REFERENCES

---

i PSA is sometimes referred to as Quantified Risk Assessment (QRA) or Probabilistic Risk Assessment (PRA).

ii. Bastin, Barón, J., Núñez McLeod, J., Rivera, S. and Perera, J. "Probabilistic Safety Assessment of the Australian Replacement Research Reactor, Risk 2004 National Conference, 8-10 November, 2004 Melbourne, Paper 215.

iii. Corran, E., Bastin, S., and Witt, H., "Reliability Analysis of a wellhead shutdown system", Commercial report ANSTO/C-140, 1990.

iv. Bastin, S., and Witt, H., "Reliability models for overpressure protection systems on offshore platforms", Commercial report ANSTO/C-230, 1991.

v. Bastin, S., "Determination of Proof Test Periods of Instrumentation in the Combined Safety System of an Offshore Platform by Risk Criteria", Commercial report ANSTO/C-346, 1994.

vi OPAL is the Open Pool Australian Light-water reactor that is currently undergoing commissioning at the ANSTO Lucas Heights Science and Technology Centre.

vii. INVAP, Probabilistic Safety Assessment (PSA), ANSTO Replacement Research Reactor Project, Prepared by INVAP for ANSTO, RRRP-7225-3BEIN-002-REVA, November 2004

viii. Barón, J., Núñez, J. and Rivera, S., “A Level III PSA for the Inherently Safe CAREM-25 Nuclear Power Station”, Probabilistic Safety Assessment and Management 96, Cacciabue & Papazoglou (Eds.), 1996 Springer-Verlag, London, ISBN 3-540-76051-2, pp. 553-558.

ix. US-NRC, (1983), “PRA Procedures Guide”, NUREG/CR-2300, Nuclear Regulatory Commission, USA, 1983, pp. 3-21.

x. IAEA, (1989), “Application of Probabilistic Safety Assessment to Research Reactors”, -TECDOC-517, International Atomic Energy Agency, Vienna, 1989.

xi. IAEA, (1995), “Treatment of External Hazards in Probabilistic Safety Assessment for Nuclear Power Plants”, Safety Series N° 50-P-7, International Atomic Energy Agency, Vienna, 1995.

xii. INEL, (2003)“Systems Analysis Programs for Hands-on Integrated Reliability Evaluations (SAPHIRE) Version 6.54”, Idaho National Engineering Laboratory, Lockheed Martin Idaho Technologies Company, Inc., USA, Fall, 2003

xiii. IAEA, (1997), “Generic Component Reliability data for Research Reactor PSA”, TECDOC-930, International Atomic Energy Agency, Vienna, 1997.

xiv. IAEA, (1988), “Component Reliability Data for use in Probabilistic Safety Assessment”, TECDOC-478, International Atomic Energy Agency, Vienna, 1988

xv. Barón, J., Núñez, J. and Rivera, S., (2002) "Human Reliability Analyses to consider Common Cause Failures for a Research Reactor during its Design Phase", Probabilistic Safety Assessment and Management PSAM-VI, Bonano et al. (Eds.), 2002 Elsevier, UK, ISBN 0-08-044121-1.

xvi. IAEA, (1993) “Probabilistic Safety Assessment for Seismic Events”, TECDOC-724, International Atomic Energy Agency, Vienna 1993.

xvii. IGNS, (1999), “Seismic Hazards Analysis - Lucas Heights, site of the High Flux Australian Reactor”, IGNS, December 1999.

xviii. Stirling, MW, and Berryman KR, (2001) “Extension to Lucas Heights Probabilistic Seismic Hazard Assessment”, IGNS, July 2001.

xix. ARPANSA, (2001), “Regulatory Assessment Principles for Controlled Facilities”, RB-STD-42-00 rev 1, Australian Radiation Protection and Nuclear Safety Agency, Regulatory Branch, October 2001

xx. USNRC, (1994), “Handbook of Methods for Risk-Based Analyses of Technical Specifications”, US Nuclear Regulatory Commission, NUREG/CR-6141 1994.

IMPROVING MAINTENANCE WORKFLOWS THROUGH THE USE OF ACTIVITY SAMPLING TO IDENTIFY AND REMOVE BARRIERS TO PRODUCTIVITY

Steve Berquist

Fluor Operations & Maintenance, Brisbane, Queensland 4064, Australia.

Abstract: The way we design and implement our maintenance management processes directly impacts the productivity we can achieve from our maintenance workforce. Activity sampling is a method that can be used to directly measure the direct utilisation achieved in the field by the craft labour and more importantly help identify the delays and support activities that are preventing craft from "getting on with the job". This information is used to target specific changes that will remove barriers to improving both wrench time and overall maintenance workforce productivity. Benefits include reduced cost and increased plant uptime. This paper discusses the elements that make up overall craft productivity and describes Australian experience in the application of activity sampling in the heavy industrial maintenance environment in both union and non union workplaces.

Key Words: activity sampling, wrench time, workforce productivity, maintenance workflow

1 INTRODUCTION

One of the accountabilities of maintenance and asset management professionals is the identification of what is the right work to be done in order to achieve the business objectives from the physical assets under management. Once the "what" has been established it then becomes necessary to establish the "how"? This is sometimes expressed as …..doing the right things and then doing those things right. The maintenance workforce is usually one of the larger groups at a plant that are expected to do these things right and attracts considerable supervision and management attention when this is not believed to be the case.

In Australia today, as in may other places in the world, the availability of skilled maintenance personnel is insufficient to match the demand for these resources in the market. Many asset owners, and contract service providers, are finding that their ability to grow their business is being restricted by the availability of skilled personnel, and the ability to leverage their workforce to support increased production output. It was reported by the Reserve Bank of Australia [1] recently that, for the first time in the history of NAB business surveys in Australia, the most popular response to the question of what factors are significantly constraining their output was 'lack of suitable labour' and not 'sales and orders'.

Accompanying this reporting of high levels of labour scarcity is the current strong Australian labour market with unemployment rates down from a peak of nearly 11 per cent in 1993 to an average of just over 5 per cent in 2005. State-wide shortages are reported in the engineering trades including metal fitters, electricians, electrical powerline trades and welders in New South Wales, Victoria, Queensland, Western Australia and others [2]. These are skills that are required in significant numbers to maintain and improve the heavy industrial assets of many businesses in this country.

In addition to a shortage of maintenance and construction personnel to support growth there is the relentless pressure on plant managers to remain in business by operating their assets at a competitive unit cost position in the global market place. Companies who can not reliably produce their product at a competitive unit cost eventually go out of business. The cost of maintenance is usually a significant contributor to the unit cost of production. It typically ranges from less than 3 per cent to more than 8 per cent of sales revenue [3] depending on the industry and much management attention is applied to reducing, or in more enlightened companies optimising, this cost.

Labour makes up a significant component of the cost of maintenance and this paper describes an approach to assisting maintenance workforces increase their productivity and thereby accomplish either the same quanitity of the "right" work for less labour cost or alternatively achieve more work for the same labour cost. Another benefit to businesses, often far exceeding the labour cost benefit of improved maintenance productivity, is the increased revenue due to less plant down time to achieve a given work scope. This is especially important during plant shutdowns or turnarounds when major sections of a plants productive capacity are taken off-line for several weeks or months. Significant revenue benefits can be achieved if the duration of the shutdown can be reduced, with revenues in the millions of dollars per day not uncommon. Conversley significant contractual penalties can be incurred if planned shutdown durations are exceeded. Finally, if labour is in scarce supply then it becomes even more of an imperative to maximise the work that is achieved from the workforce that one already has.

2 CRAFT PRODUCTIVITY

There are three basic elements that together define overall workforce craft productivity. These are

a) Direct Utilisation: the percent of time that craftspeople are performing direct work i.e. wrench time

b) Direct Productivity: the actual units of work completed per unit time expressed as a percentage of a defined standard for that work

c) Rework: the percentage of work performed that must be redone to correct errors

When these three elements are combined together craft productivity becomes an overall measure of the amount of labour required to perform direct work compared to that required to perform the same direct work without any losses in efficiency or effectiveness. It is the amount of work that actually gets done per paid hour.

$$= \quad \times \quad \times (\quad - \quad)$$

where: DU = direct utilisation DP = direct productivity RW = rework

Each of the three elements defined above will be examined in more detail below.

2.1 Direct Utilisation

Direct utilisation is the ratio of actual direct activity measured to the target direct activity for our workforce. It indicates how our actual direct activity compares to our desired direct activity. Fluor sets 50 per cent as the direct activity target for a maintenance workforce.

The formula to calculate direct utilisation is below.

$$= \frac{\quad}{\quad}$$

Direct activity is the percentage of time that a maintenance technician spends on activities that are directly advancing the execution of a maintenance task. This is often referred to as “wrench time”, loosely illustrating that the technician actually has a “wrench” in hand performing direct activity advancing the job towards completion i.e. welding, grinding, painting, bolting, measuring, fitting, rigging etc. Conversely, indirect activities are those things that are not directly advancing the job. If these activities could be reduced or eliminated, more time would be spent by the maintenance technician “on the job”. Some examples of indirect activities are travel to and from the job, waiting for instructions, waiting for permits, waiting for materials, getting materials from the store, planning the job before starting to work, taking rest breaks and meal breaks.

While not all types of indirect activities are undesirable it is important to acknowledge that this category of activity exists. When measures are taken to reduce the amount of time in the day a technician spends on indirect activity it will create the opportunity for more direct activity and hence for more work to be accomplished.

Fluor, in Australia, finds organisations that have not implemented best practices for maintenance planning and scheduling typically have direct activity in the 20 per cent to 30 per cent range. World class performance can result in direct activity of 50 per cent or more.

A measured direct activity (wrench time) of 30 percent equals 60 per cent direct utilisation i.e. the maintenance work processes on site are resulting in people working at only 60% of the Fluor target for direct activity i.e

$$= \frac{\quad}{\quad}$$

2.2 Direct Productivity

This is the percentage of time actually taken to perform a task versus the standard time established for that task. Standard times are established for tasks based on the time an average skilled technician will take to do the job working at normal pace using a prescribed method without breaks. Standards are more likely to have been established by planners for frequently recurring maintenance tasks than for other types of work. Where standards do not exist it is possible to simply use the estimated time assigned by the planner for the task as a benchmark for direct productivity calculations.

The Gulf Coast of Texas is a well know benchmark used by construction industry estimators for comparing labour productivity in other locations to that experienced from workers on the Gulf Coast. Once the productivity factor is known, labour estimates can be then be factored for other locations of the world.

2.3 Rework

This is the percentage of craft labour that is required to perform a task a second time because of errors with the performance of the task the first time. Rework has many causes and is usually a topic of heated debate at all levels in the maintenance organisation. Few organisations measure and record rework.

It is possible to use a computerised maintenance management system to check for occurrences of repeat maintenance intervention on equipment within a defined time period (often one month is selected) from the last work order being completed. This can be used as a rework metric as follows

$$= \frac{Total\ Hours\ of\ Repeat\ Maintenance\ Intervention\ Work}{Total\ Maintenance\ Hours\ Worked}$$

The target for rework is zero. The measurement of rework generates focus on this problem and will immediately result in the desire to categorise the causes for this loss. Causes include faulty materials, poor procedures, lack of training, lack of time, rushing the job, inadequate supervision or incorrect work scope. Once the causes are understood corrective actions can be taken.

3 DRIVERS OF CRAFT PRODUCTIVITY

Craft productivity is a lagging indicator of the effectiveness of the maintenance management processes implemented at a plant. Where plant management has developed and implemented effective integrated work processes, together with providing effective training, supervision, motivation and support to the workforce, craft productivity will be significantly higher than where these factors are absent. Examples of processes that have a significant impact on craft productivity include

- planning
- scheduling
- permit to work
- materials management
- contractor management
- shift handover
- site logistics
- job safety
- technical support

The challenge to craft productivity improvement is to first measure what the current level of productivity is and then to identify what current barriers exist preventing productivity from increasing. In Fluor's experience, the craft workforce understands what the barriers are that hinder them from "getting on with the job". If asked they will tell you. Most barriers are sources of frustration to the workforce and they will gladly point out what should be changed! It is management's responsibility to listen and take those actions necessary to address barriers in a prioritised manner.

4 ACTIVITY SAMPLING

Activity sampling is the process of measuring an activity of a population by the acquisition of sufficient random samples to make a statistically valid conclusion about that activity for the entire population. Activity sampling, often called work sampling, is a recognised industrial engineering process and has been used by Fluor for over 20 years to measure the labour productivity of its construction workforces and for more than 15 years for maintenance workforces.

The activity sampling process is one where an observer makes a snapshot observation of a person or group of people, records the observation immediately and briskly moves on until the next observation is made. Observations are made by walking a pre-determined route around the plant. Each observation should take no longer than a few seconds to capture. The workers being observed are not subject to continuous surveillance, so are more accepting of this style of observation than if they were placed under constant observation.

Activity sampling is a practical alternative to the continuous observation method, sometimes known as time and motion studies, where a person's activity is observed continuously over a period of time. The continuous observation of a person is usually unwelcome in most workplaces in Australasia and will quickly lead to objections from the workforce and an immediate change in their normal behaviour. Industrial action is a strong possibility in many Australian workplaces if a workforce is subjected to a continuous observation process. An exception to this is direct observation performed as part of a safe work observation process where the objective is to improve people's safety at work.

4.1 Categorisation of Work

The three major categories of work used in work sampling of a maintenance workforce are Direct Activity, Support Activity and Delays. These are defined as follows:

a) Direct Activity: Performing those elements of a task that directly advance its completion. Examples include:

- Working with tool in hand
- Measuring up prior to cutting
- Using a test instrument
- Directing a crane lift
- Operating a forklift
- Erecting a scaffold

b) Support Activity: Performing those elements of a task that indirectly advance its completion. Examples include:

- Travel to get/with equipment
- Travel to get/with materials
- Planning on the job
- General travel
- Checking/receiving a work permit
- Daily tool box meeting

c) Delays: Periods of lost time that prevent Direct or Support work. Examples include:

- Waiting for a work mate
- Waiting for a vehicle
- Waiting for materials
- Waiting for instructions
- Waiting for a permit
- Taking a smoke break

The immediate work area is defined as being with 3 to 4 metres of the task. Most activity occurring in the immediate work area is considered direct unless it is a delay. For example moving with materials in the immediate work area is a direct activity, whereas moving with materials outside the work area is a support activity and it could be questioned why the journey was necessary.

Support activity and delays are losses that reduce direct activity in the field. These losses can be considered as one or other of following:

- Inherent Losses: These are the minimum requirements of time related to travel and logistics assuming pre-planning and preparations for work is complete and without errors e.g. the need to travel from the workshop out to the job site at the start of the day or the need to travel to and from the lunch room.
- Planning Losses: These are all other time requirements beyond direct utilisation and inherent losses e.g. the need to return to the store from the job site to get spare parts that were not identified on the job plan as being required or the need to wait until another crew moves out of the way before starting on a new job.

Inherent losses will occur even with the best planning and these activities can never be eliminated, only minimised. Planning losses are activities that are possible to eliminate given better planning. The aim of the maintenance activity sampling study is to increase craft productivity by reduction of the inherent and the planning losses.

4.2 Statistical Basis of Activity Sampling

The statistical basis of activity sampling requires sampling performed in an unbiased way. Randomisation is used to eliminate bias. Examples of the factors to be randomised include; starting time for each sampling activity, direction of movement along each sampling route and the location for starting the sampling along each route. Because the very act of taking an observation can change the behaviour of the population being sampled it is important to take enough samples over a number of days to ensure the final results are truly representative of the population's actual direct activity.

Using the assumption that a population's direct activity follows a normal distribution, the formula used to solve for the minimum number of required sample observations is [4]:

$$n = \frac{z^2 p^\wedge (1 - p^\wedge)}{E^2}$$

where: n = number of sample observations required

z = standardised normal random variable corresponding to the level of confidence in result i.e. z = 1.96 for a 95% confidence level in our result.

$p^\wedge$ = an estimate of the populations proportion i.e. estimate of the direct activity expected

E = one half of the total precision range (error) acceptable i.e. E = +/- 0.025 for a 5% error

The variable $\hat{p}$ is usually estimated prior to sampling at a value of 0.5 i.e. 50 percent direct activity.

Utilising the values suggested above, a total of n = 1537 observations must be made of either the total workforce, a plant area or a workgroup, in order to be 95 percent confident that the direct activity sampled does in fact correspond to that of the workforce population being sampled with a precision range (error) of 5% i.e. +/-2.5% .

If the actual direct activity $\hat{p}$ for the workforce is found to fall above or below 0.5 during the course of the sampling period, the minimum value for n can be adjusted by iteratively calculating its value as the workforce $\hat{p}$ is sampled.

4.3 Route Design

The design of the routes used for activity sampling of a maintenance workforce requires consideration of the following factors

- The sites overall layout. The larger the site (horizontal and vertical) the more routes will be required.
- The expected location of maintenance workers around the site – plant areas, office buildings, workshops, stores and break areas. The routes combined need to cover 90 percent to 110 percent of the workforce each time the plant is walked. Some people may miss being observed, others may be observed twice.
- The presence of hazards on site and the ability for observers to remain safe while on their route. Some areas may need to be excluded from a route e.g. electrical switch rooms, confined spaces etc.
- The number of observers available to assign. The fewer observers then the fewer routes are possible.
- The location of, and activity on, main thoroughfares and roads. May require splitting across several routes.
- Each route to take a similar time to cover

4.4 Route Schedules

The schedule for covering the plant is developed based on the following

- The shift work pattern to be sampled i.e. day work only, 24 hour shift, weekdays only etc.
- The start and finish time for the shifts being covered.
- Random starting times for each round of routes. All routes walked together starting at the same time.
- The timing of the allowed breaks taken by the workforce i.e. tea breaks and lunchtime.
- The direction the route will be walked i.e. forwards or reverse direction
- The start position along the route that the route will be walked from. Route can be started from a mid point, start or finish location.
- The time taken to walk the route. The longer it takes the fewer routes can be scheduled per day.

The aim is to have sufficient routes walked per day to collect the required number of observations over shortest number of days commensurate with covering the shift patterns of the workforce being studied. For example it might be required to cover every day of the week and every hour of the day between 7am and 7pm when studying a maintenance workforce that works days only but has coverage on site every day of the week. If there are a significant number of maintenance people working night shift then it may be an objective of the study to sample nightshift work also.

Once the routes have been designed and the schedule established it is necessary to then prepare both the workforce and the observer team for the work sampling to commence.

5 CRITICAL PREPARATION FOR SUCCESSFUL ACTIVITY SAMPLING

There is significant preparatory work required in order to ensure activity sampling can be performed successfully in a heavy industrial workplace. The key activities required to ensure success are discussed briefly below.

5.1 Management Commitment and Alignment

Unless the plants site management are committed to improving craft productivity and clearly understand how the activity sampling process will contribute to that goal, the probability of success will be reduced. It is management that will be required to release personnel to participate in the activity sampling team and to support the team through assisting to eliminate barriers that can arise during the course of the activity sampling process. Common barriers are those associated with releasing personnel from normal duties and implementing visual identification to enable various work groups to be accurately identified by the observers.

5.2 Communication

It is critical that the workforce being sampled is engaged and supportive of the sampling process. Fluor's experience is that Australian workforces are generally accepting of most initiatives genuinely aimed at improving their workplace experience. This includes anything that will make it easier for them to "get on with the job". The identification and elimination of barriers to getting work done is usually welcomed and the workforce views the activity sampling exercise as an opportunity to validate with management their previously expressed views and ideas for improving the workplace.

Fluor communicates the activity sampling process to the workforce either through a presentation direct to the workforce or to representatives of the workforce. The presentation explains the benefits the workforce can typically expect from implementing solutions to its findings. Concerns that the workforce commonly have are reduced through the following means.

- Request workforce representation on the observation team.
- Advice that supervisors are not included in the observation team.
- Advice that people's names are not recorded and also that the job observed is not recorded.
- Advice that full training is provided to all observers.
- Make all observation record sheets available to the workforce to review at any time. Invite them to look immediately after making an observation in the field especially during the first few observation rounds.
- Advice that a closeout presentation will be provided to the workforce and their input sought in improvement planning.
- Advice that the opportunity exists to identify to management current barriers to improving productivity e.g. tools, vehicles, workshops, stores etc.

5.3 Observer Team

The observer team must be fully trained in order to ensure that accurate observations are made and that there is no inherent bias in the observation process. It is the responsibility of the activity sampling team leader to identify and correct individual observers who are making suspect observations. The observer training must include the following:

- An understanding of the three elements of craft productivity.
- The ability to identify and categorise members of each of the workgroups included in the study
- The ability to identify direct activity from the various categories of delay activities and support activities.
- The ability to question people respectfully in order to correctly categorise the activity first observed.
- The ability to identify the reason why delay or support activities observed are occurring.
- Knowledge of each route that is to be walked.
- A test to confirm that each student has been successfully trained.
- Field practice, to confirm to the leader that each team member is ready to deploy.

The observers may operate in the field either singly or in pairs depending on numbers available. More than two people is not advisable as it makes the observers too conspicuous and therefore enables people to change their behaviour when the observers approach. The advantage of working in pairs is that observer safety is enhanced out in the plant and that discussion on observations and reasons for the activities observed can be held. It also helps maintain observer motivation and allows observers to swap partners occasionally.

5.4 Diagnostics

It is critical that observers are able to not just correctly categorise and record the work activity observed but to also attempt to determine the cause of the activity that was observed, where it is not readily obvious. By asking the people what caused the need to travel to the store or what resulted in the delay while they wait for a colleague to return from the workshop with a tool, it becomes possible to begin to identify where the work processes used on site are failing to support craft productivity. If all that results from an activity sampling study are values for percent direct activity and percentages of delay and support activities, much of the value of the exercise is lost. Activity sampling of a maintenance workforce should provide actionable information that will help drive improvement in craft productivity, not simply state what the current wrench time is.

6 AUSTRALIAN EXPERIENCE APPLYING ACTIVITY SAMPLING

The authors company, Fluor, has successfully implemented labour activity sampling of maintenance workforces in three states of Australia during the past year – Western Australia, Victoria and Queensland. These workplaces include workforces employed under the following industrial instruments; Australian Workplace Agreements (non union collective agreement), Enterprise Bargaining Agreements (union collective agreement) and Individual Contracts of Employment (common law contract).

As stated earlier in this paper the key to acceptance of the workforce in each case to the application of the activity sampling methodology has been the genuine commitment of management to improving the workplace experience of maintainers through the elimination of those frustrating barriers and delays experienced by people every day. When the workforce understand and accept that the process is not about headcount reduction but rather about improving the maintenance workflow process they are usually willing to accept it and contribute as members of the activity sampling team.

In the authors experience the workforce sees the activity sampling process as providing them opportunities to be heard by management and to prove through the collection of hard data that the problems they have been complaining about for years are real issues that management should address. This objective aligns with management's objective to improve craft productivity enabling more direct work to be accomplished (maintenance work and/or improvement work) by the workforce. Business benefits include reducing the maintenance backlog, reducing the amount of overtime required, reducing the requirement for contractor labour, improved equipment reliability through improved preventive maintenance program compliance and the ability to release craft from the tools to work on improvement projects e.g. RCFA team participation, review of PM's, review of BOM's etc.

At all times it is necessary to acknowledge to both the workforce being asked to undergo the observation of their work activity and to their supervision, that the activity sampling process is somewhat of an intrusion into their privacy, that their patience and tolerance of having people somewhat "interrogate" them over what they are doing is appreciated and that the observations will come to an end shortly. The supervisors are often more concerned than the maintenance workforce because the results may reflect unfavourably upon their supervisory performance.

Fluor's experience in Australia includes successful introduction of labour activity sampling to our client's own in-house craft employees, Fluor's own craft employees and other contractor's employees. Both permanent (full-time) and casual (temporary hire) employees have been successfully engaged. Regardless of employer, it is important to include all craft in the communication process.

7 SOME AUSTRALIAN MAINTENANCE RESULTS

Table 1 contains examples of results of labour activity sampling of maintenance from three Australian worksites

Table 1

Summary results from three Australian workplaces

Site	Industry	State	Direct activity %	Support activity %	Delays %
1	Mining/Mineral Processing	Western Australia	33.9	31.8	34.3
2	Power	Victoria	27.6	32.1	40.3
3	Mineral Processing	Queensland	25.5	40.1	34.4

These results are typical of sites where best practice maintenance work processes have not been implemented. It is clear that the direct activity measured is significantly lower than the target of 50 per cent. The opportunity exists in all these sites to reduce time spent by the craft workforces on delay and support activities and increase time spent in direct activity. Figure 1 below illustrates this potential.

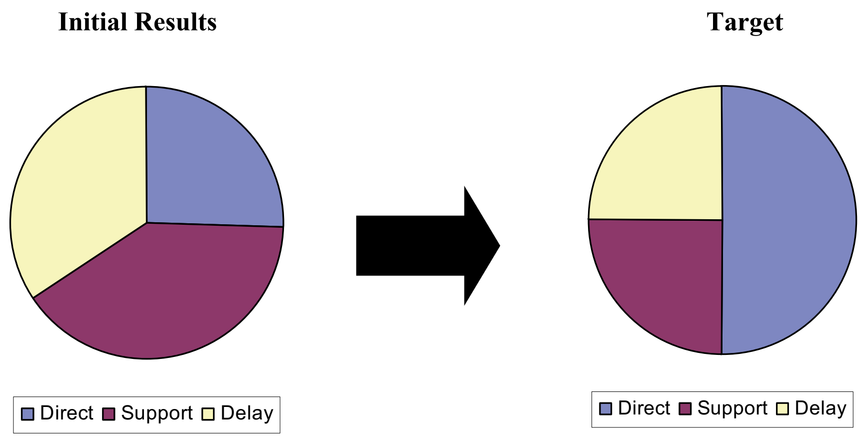


Figure 1: Typical activity sampling results and comparison to best practice targets.

As an example, if the direct activity of site 3 can be increased from 25.5 per cent to the target of 50 per cent, that workforce could achieve double the maintenance workload that it previously was capable of. This improvement would only occur following changes successfully implemented to address the causes of the excessive delay and support activities measured.

Figure 2 below shows the major causes of the delays and support activities measured at one of the three Australian sites listed in Table 1.

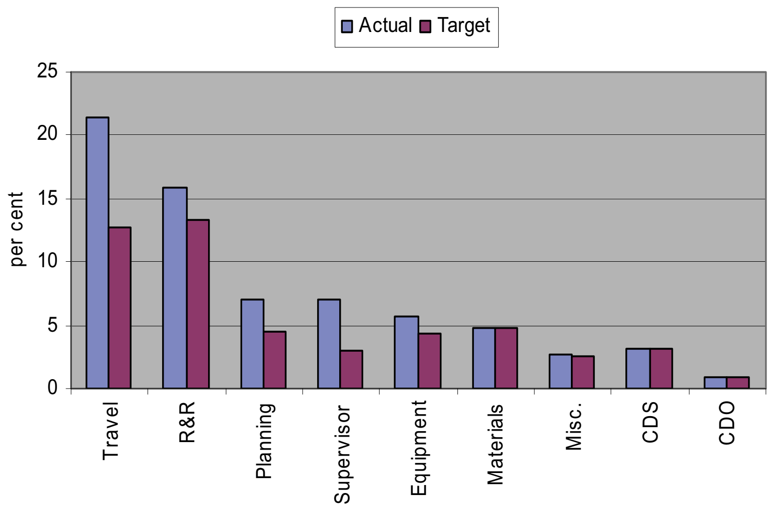


Figure 2: Graph of major reasons for losses in direct activity for an example Australian maintenance site.

The graph shows that the biggest opportunities at this site lie with reducing the time spent on general travel around the site, in this case from 21.4 per cent down to 12.7 percent. The second largest opportunity lies in reducing the time spent on supervisory related delays from 7 per cent down to 3 per cent. Planning is the third area of opportunity for this site where craft spend time on planning related activities that should have been performed by others.

It is the task of the activity sampling team to identify the gaps between measured activities and best practice targets for each activity. Fluor uses its experience from over 20 years of performing activity sampling on both maintenance and construction sites to establish achievable targets that can be met through the application of best practices in each loss category.

Other forms of analysis performed that provide additional insight into the causes of losses in direct activity are the review of results by any of the following; shift, day, hour, crew, route, plant area, contractor or trade. When subsequent studies are performed it is also possible to compare results by time of year to assess the impact of climatic conditions on direct activity.

7.1 Improvement Planning

Improvement planning is the next step following on from the establishment of the current situation by conducting an activity sampling study. Improvement planning is most effective when the craft workforce is included in the planning process. Maintenance best practices in each of the key loss categories must be identified and utilised by the improvement planning team in the redesign and planning for implementation of changed work practices in order to significantly improve direct activity to sustainable target levels.

7.2 Implementation

Sustainable improvement only occurs after changes are made to the work processes in place at the time of the initial activity sampling study. There may be an initial improvement in direct activity simply through drawing the workforce's attention to the fact that the current wrench time is low and what some of the major causes of losses are. For example, there may be an initial improvement in some areas such as early stop and late start to work around scheduled breaks. Unless there are changes to things such as the way work is allocated to individuals or the way craft are briefed by supervision any initial improvement will soon be lost.

Table 2 shows an example of the improvements made by an Australia craft workforce through the design, planning and implementation of a wide range of maintenance best practices. The results are shown for three cycles of activity sampling studies over a period of 12 months.

Table 2

Summary results from three cycles of maintenance labour activity sampling over a 12 month period

	June 2004	April 2005	July 2005
Direct activity %	33.4	36.4	41.4
Support activity %	30.8	29.6	29.2
Delays %	35.8	34.0	29.4

8 CONCLUSION

Activity sampling of maintenance work is an effective method for establishing the level of craft direct utilisation or wrench time. Direct utilisation is one of the three elements of overall craft productivity, the others being direct productivity and rework.

Activity sampling has been successfully applied by Fluor in maintenance workplaces in heavy industries around the world including three states in Australia. It has been acceptable to both union and non union craft employees when communicated well and when supported by management's commitment to remove the barriers and frustration that get in the way of craft simply getting on with the job.

Improvement in direct utilisation will occur when the gaps between best practice and current practice are identified in a number of maintenance areas identified from the activity sampling as the main causes of loss in direct productivity. The craft workforce must be included in the improvement planning process.

9 REFERENCES

[1] Edey M. (2006) The Australian Economy – Prospects for 2006 and Beyond. Address to the 2006 CEDA/Fujitsu Economic & Political Overview Sydney 24 February 2006.

[2] Australian Government Department of Employment and Workplace Relations (DEWR). Skills in Demand Lists States & Territories – 2005.

[3] Fluor's Best of the Best Maintenance Benchmarks. Internal publication.

[4] High Level Manpower Activity Analysis. Fluor Operations & Maintenance confidential internal document.

[5] Finigan T J, Fluor (2005) Boost Maintenance Staff Efficiency. Article in Chemical Processing magazine, October 2005

[6] DCAA Contract Audit Manual 7640.1 – January 2005. Defence Contract Audit Agency (DCAA).

Acknowledgments

The author acknowledges the colleagues, clients, contractors and craft that have cooperated successfully on the activity sampling studies referred to in this paper. Support is gratefully acknowledged from Mr Brian Fox of Fluor's Greenville South Carolina office.

FAULT DIAGNOSIS OF CARBON FIBRE COMPOSITE MASTS

Dabrowski Z.[a], Dziurdz J.[a] and Skorski W.[b]

[a] *Institute of Machine Design Fundamentals, Warsaw University of Technology, Narbutta 84, 02-524 Warsaw, Poland*

[b] *Institute of Aeronautics and Applied Mechanics, Warsaw University of Technology, Nowowiejska 24,00-665 Warsaw, Poland*

Abstract: More often in the world, in many applications, composite materials replace traditionally applied materials like steel, aluminium alloys etc. Especially they have an application when the product must be light and should have specific mechanical and dynamic features, important for the sake of place of application and type of work. One of this application is the mast of high-performance sail boat. Applied materials in process of its production are expensive, and technology and process engineering are complicated and time-consuming. Mast, for the sake of characteristic kind of work and because it is very important constructional element, should have very big reliability. Other important thing is ability to early defect detection of reinforcement structure at the stage of production and then in the middle of operational use. This ability has large influence on safety of deck crew. One of the most often and invisible defect is enlarging of holes around rivets and bolts fasten mounting elements of rigging caused by strong dynamic overloads. In this paper Authors made an attempt to finding this kind of defects by vibroacoustic techniques and try to determine proper symptoms. Authors postulate that change of hole diameter should change dynamic response (transmittance) of a system on dynamic coercion.

Keywords: carbon fibre composite mast, dynamic response, coherence function.

1 INTRODUCTION

In the contemporary world in many different areas composite materials replace those traditionally used: steel, aluminium, etc. It happens especially when the product should possess small mass together with particular mechanic and dynamic characteristics, important because of its position and character of work. One of such applications is the mast of a sport sailing boat. The materials it is produced of (e.g. carbon and Kevlar fibres) are expensive, while technologies of production applied tend to be complex and time-consuming, and consequently expensive as well.

The life-span of the modern rigging of racing yachts, which is designed with minimum safety factors, is necessarily limited. The technology of rigging fastening itself results in local brakes of reinforcing fibers, or the basic supporting structure, in spite of variety of reinforcing and stiffening solutions applied; all these results in important construction difficulties. Enlarging mast's openings are one of the principal, invisible damages, followed by delamination around bolts and pivots of fittings. The first structural degradation of the composite (e.g. delamination) happens already during its mechanical processing. It influences further development of the composite's delamination, finally resulting in lost of its mechanical characteristics, happening even earlier that it would be forecasted taking the analyses of tolerable stress. Initially, it is only a phenomenon of local dimension but in consequence it results in weakening the whole structure of the construction. Nowadays both monitoring and detecting damages of the composite, as well as identifying its state is carried out applying non-destructive methods, mainly roentgen graphic and ultrasonic ones, and only recently some tests using vibroacoustic diagnosing have been done.

The vibroacoustic method has the advantage of allowing localization of any discontinuities existing inside of the material (it is clearly visible in the pictures of samples: darker spots between openings). The roentgen graphic method is limited only to the immediate surroundings of openings and edges. However, it is much more accurate (to achieve approximately similar resolution in ultrasonic scanning, it would have to take a very long time: assuming a step of 0.1mm scanning of one sample would take about 75 hours (let's note that the resolution of film is higher anyway).

The fact that correct diagnosing is possible using both methods doesn't change the truth that these are only laboratory methods that can not be applied directly to an object. Therefore, the aim of the Authors' studies was to answer the question whether it is possible to find a simple vibroacoustic symptom sensible to the increase of discontinuity of the composite structure.

2 BRIEF THEORETICAL EXPLANATION

The vibration analysis of composite masts of non-circular section has been treated by authors [5,6]. The analysis clearly show, as it has been expected, that even with the small amplitudes the non-linear effects of the system can be detected.

However, in spite of all the work done in this area of studies, there is still no mathematic (digital or analytical) model available, accurate enough as to let the analysis of form variation of vibrations accompanying propagation of a small structural damage. On the other hand, we know that the initial stages of damages are so soft in terms of energy that in practice we are not able to observe any change in general vibration level. In these circumstances the only thong to do is to make use of he fact that the proportion of damage changes the non-linear disturbances in a system. In [6] the authors applying the analysis of non-linear differential equations have shown that slightly non-linear disturbances can be observed in spectral characteristics of the system even though the medium value of spectrum (general vibration level) does not change. The persisting problem consists of searching for measures that would be sensible to this type of disturbances. One of them is the function of common coherence (the condition is that the system of analysis should be of appropriate accuracy). Let us consider a simple model shown in the picture 1.

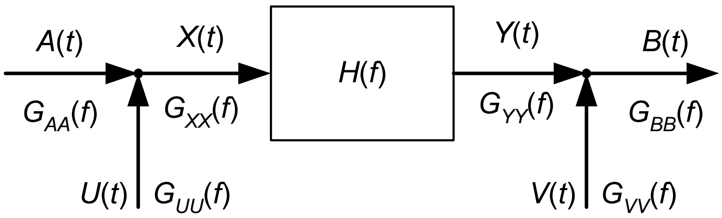


Figure 1. Model of the analyzed system.

When the input $U(t)$ and output $V(t)$ disturbances equal 0, the transmittance values are deduced from the generally known relationships

$$H_1(f) = \frac{G_{AB}(f)}{G_{AA}(f)} \qquad (1)$$

$$H_2(f) = \frac{G_{BB}(f)}{G_{BA}(f)} \qquad (2)$$

here:

$G_{AA}(f) = \left(G_{AA\,\mathrm{RMS}}(f)\right)^2$	averaging power spectrum of the input signal,
$G_{BB}(f) = \left(G_{BB\,\mathrm{RMS}}(f)\right)^2$	averaging power spectrum of the response signal,
$G_{AB}(f)$	averaging mutual spectrum of the signal at the system's input and output (the spectrum correlated with $G_{BA}(f)$),
$G_{BA}(f)$	averaging mutual spectrum of the signal at the system's input and output (the spectrum correlated with $G_{AB}(f)$).
$G_{UU}(f)$	averaging power spectrum of the input disturbance,
$G_{VV}(f)$	averaging power spectrum of the output disturbance,
$G_{XX}(f)$	averaging power spectrum of the non-disturbed input signal,
$G_{YY}(f)$	averaging power spectrum of the non-disturbed output signal.

They meet the condition without disturbances:

$$H_1(f) = H_2(f) \qquad (3)$$

and function of ordinary coherence equal 1:

$$\gamma_{AB}{}^{2}(f)=\frac{H_1(f)}{H_2(f)}=1 \tag{4}$$

When disturbances appear both at the input and output, the relationship ( ) can be brought to the following form:

$$\gamma_{AB}{}^{2}(f)=\frac{\gamma_{XY}{}^{2}(f)}{1+G_{UU}(f)/G_{AA}(f)+G_{VV}(f)/G_{BB}(f)+G_{UU}(f)/G_{AA}\cdot G_{VV}(f)/(f)G_{BB}(f)} \tag{5}$$

On many occasions we can assume that a non-linear system works in the same way a linear one does, when both kinds of disturbances are present, so we can bring the formula to the following form [4]:

$$\gamma_{AB}{}^{2}(f)=\frac{\gamma_{XY}{}^{2}(f)}{1+\Delta(f)}=\frac{H_1(f)}{H_2(f)}\cdot\frac{1}{1+\Delta(f)} \tag{6}$$

where $\Delta(f)$ is the correction coefficient of coherence function on account of slightly non-linear disturbances. If we assume in diagnostic applications that each state comes accompanied by different disturbance, the coefficient $\Delta(f)$ becomes a natural state parameter, and subsequently, according to the relationships (1) and (2) both transmittances (H_1 and H_2) can be such a parameter in certain frequency bands. We decided to check experimentally whether the same effect can be observed for the propagation of damage in a composite mast.

3 EXPERIMENTAL RESEARCH

Let us follow the experiment aiming at finding a way of reproducing diameter of the opening in form of vibration signal. Although it is a laboratory experiment, translating its results into a real object is relatively simple.

The measurement set used in the test (Figure 2) consisted of the following elements (all produced by Brüel&Kjaer):

- a impact hammer B&K8202 equipped with force sensor B&K8200 – used for impulse forcing of vibrations of the studied mast together with input measuring,
- piezoelectric accelerometer B&K4284 – sensor used for measuring a response for impulse input,
- preamplifier B&K2692 „Nexus" – a device adjusting the signal level to the parameters of analyzer's input,
- dual channel analyzer B&K2034 – principal element of the system, used for recording and analyzing measurement signals.

Figure 2. Elements of the measurement set.

It was assumed that positions of the vibrations sensor and the impact point of hammer should be located on the opposite ends of the studied element. The preliminary tests were done for samples of composite masts, each about 1.1 m long, with centrally situated openings (symmetrically at both sides of the mast). Positions of the accelerometer and the impact point of hammer (distance about 1 m) relative to the openings are shown in the figure 3.

The preliminary tests consisted in searching for symptoms that could point to changes of diameter of the openings in the sample of composite mast. Four measurement series for various states were made:

- for the undamaged sample,
- for the sample with an opening ø10mm approximately in the middle of the beam, on its side marked by the axis *y*,
- for the sample with two openings ø10mm approximately in the middle of the beam, on its sides marked by the axis *y* and *v*,
- for the sample with two openings ø14mm approximately in the middle of the beam, on its side marked by the axis *y* and *v*.

side marked by the axis y and v.

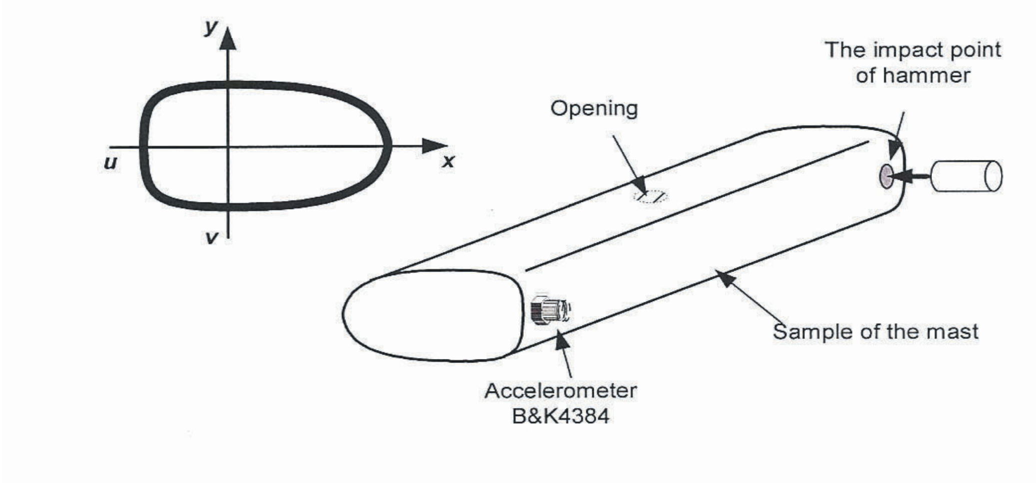


Figure 3. Positions of the accelerometer and the impact point of hammer.

Preliminary measurements and analysis of the results were carried out for several positions of the accelerometer (*u, y, v*) and the impact point of hammer (*y, u*) for three frequency ranges:

- 1600 Hz (resolution 2 Hz),
- 3200 Hz (resolution 4 Hz),
- 6400 Hz (resolution 8 Hz).

For further analyzing the frequency range of 3200 Hz was chosen, combined with positioning the vibration sensor and the impact point of hammer in axis y and u. Using the given characteristic of the coherence function which permits to define linearity or non-linearity of the system's response, the following frequency ranges were singled out in the analyzed functions:

- dominating response of linear part of the system (up to 900 Hz),
- appearing response of non-liner part of the system (above 900 Hz).

The figure 4 shows way of fixing the accelerometer to the tested samples.

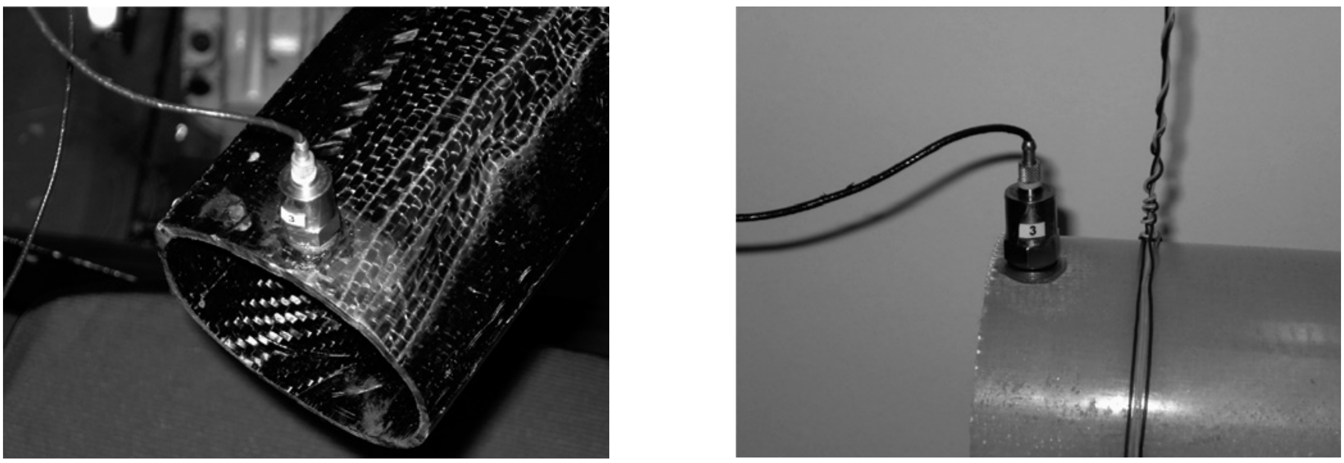


Figure 4. Fixing of the vibrations acceleration sensors.

4 ANALYSIS OF THE RESULTS

It seems obvious that for such a type of construction, strongly anizotropic, symptoms of the induced changes (damages) should be searched in the range of non-linear response of the system. In our case the following functions were registered within the capacity of the analyzer B&K2034:

- averaging spectrum of RMS values of input (force of impact of percussion hammer) G_{AA} RMS (f),
- averaging spectrum of RMS values of the system responses (vibrations acceleration at opposite side of the sample) G_{BB} RMS (f),
- frequency response of the system, admitting disturbances of the input signal $H_1(f)$
- frequency response of the system, admitting disturbances of the input signal $H_2(f)$
- coherence function between the input signal and the output signal $\gamma^2_{AB}(f)$.

All the functions serve in general for analyzing non-linear systems (or linear responses of non-linear systems). Some examples of the results of analysis are shown in the figure 5.

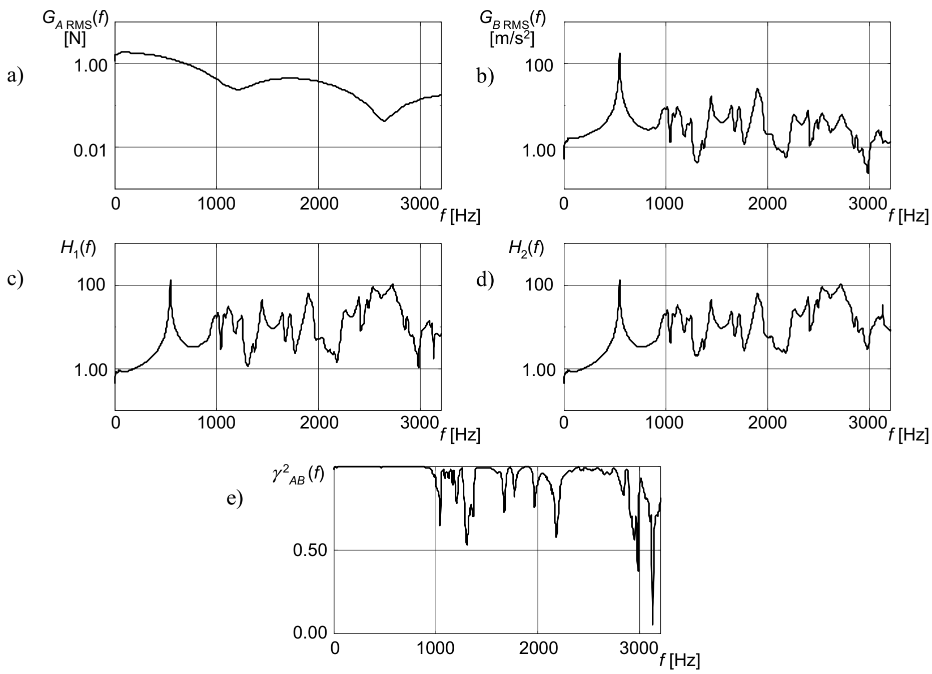


Figure 5. Visualization of measurement results (an example): a) averaging RMS spectrum of input, b) averaging RMS spectrum of responses of the system, c) frequency response of the system H_1,d) frequency response of the system H_2, e) coherence function γ^2_{AB}.

As in the tested system there are no signal disturbances at input and output of the system, the transmitantion H_1 should equal the transmitantion H_2. Therefore an ideal case should consist of a system, where coherence function between a signal originated in the source and the output signal is equal. It can be observed that value of the coherence function shown in the figure 5e is smaller than one within considerable frequency range. We should then consider reasons of such a situation. With no exterior disturbances, lack of feedbacks between the recorded signals and relatively high accuracy of measurement, the only reason seems to be non-linearity of the tested element [2, 3].

The figure 6 shows the ordinary coherence function between the input (the impulse of force coming from the hammer) and the system response (signal recorded using the accelerometer) for subsequent trials, while the figure 7 shows changes of transmitantion (system amplification factor) H_1.

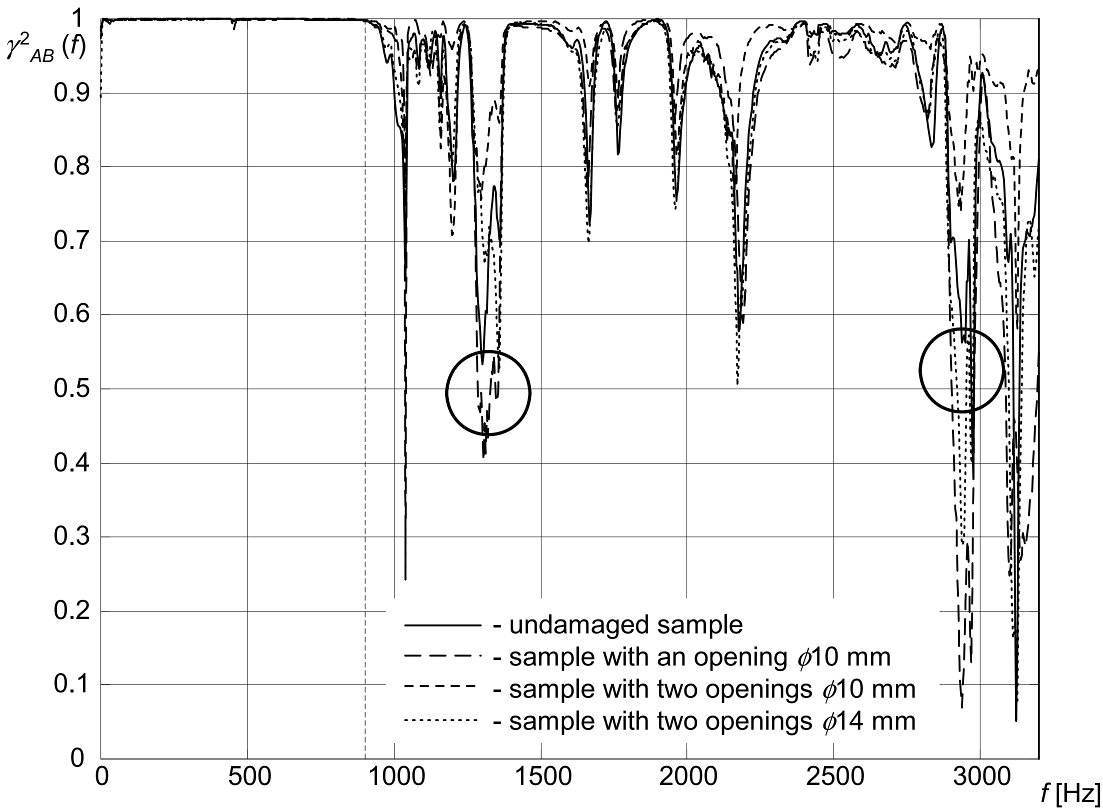


Figure 6. Values of coherence function compared for the tested samples.

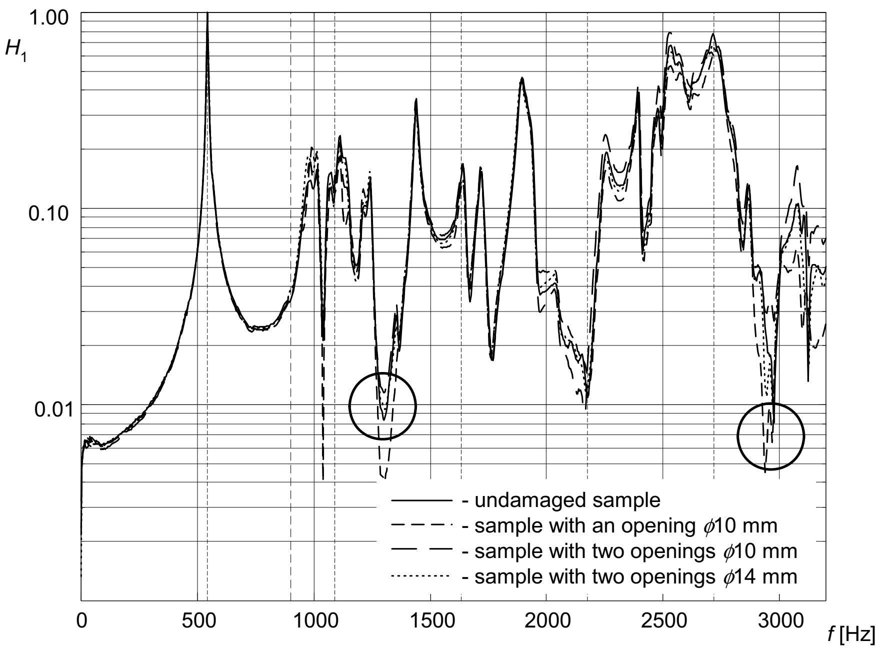


Figure 7. Values of amplification factor (transmitantion module) H_1 calculated for the tested samples.

Analyzing the results of measurements it can be said that structural disturbance of tested material (diameter of opening) gives a repeatable and intelligible picture both for the coherence function and for the amplification factors (transmitantion module).

In the main part of the research subsequent measurement series were made for different diameter openings centrally located on both sides of the mast's profile. The tests were done for openings of diameters of 4, 6, 8, 10 and 12 mm.

The measurements and analysis were carried out in the same way as in the preliminary tests. The figure 8 shows changes of value of the amplification factor (transmitantion module of the system) H_2 in frequency function for different diameters of openings drilled in the sample.

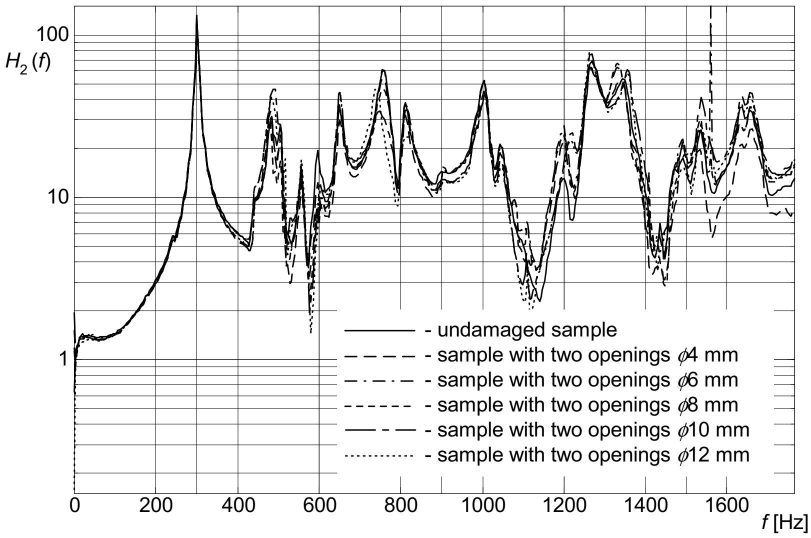


Figure 8. Changing value of the amplification factor (transmitantion module) of the system) H_2 in frequency function for different diameters of openings drilled in the mast.

Figure 9 shows values of the amplification factor (transmitantion module) of the system) H_2 in frequency function for different opening diameters in a chosen frequency band (around non-linear resonance frequency).

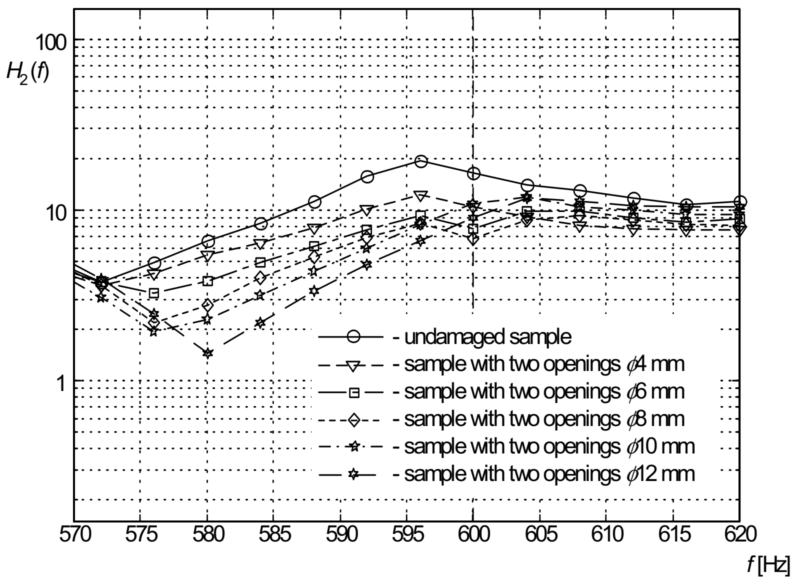


Figure 9. Changing value of the amplification factor (transmitantion module) of the system) H_2 in frequency function for different diameters of openings drilled in the mast (within the diagnostic band).

5 CONCLUSIONS

If we analyze amplification factor of the system in a chosen frequency band, we can observe a monotonic change of value (for several subsequent frequencies), depending on changes of the opening diameters. The figure 10 shows changes of value of amplification factor in function of changes of diameters of the openings in the mast (for selected frequencies).

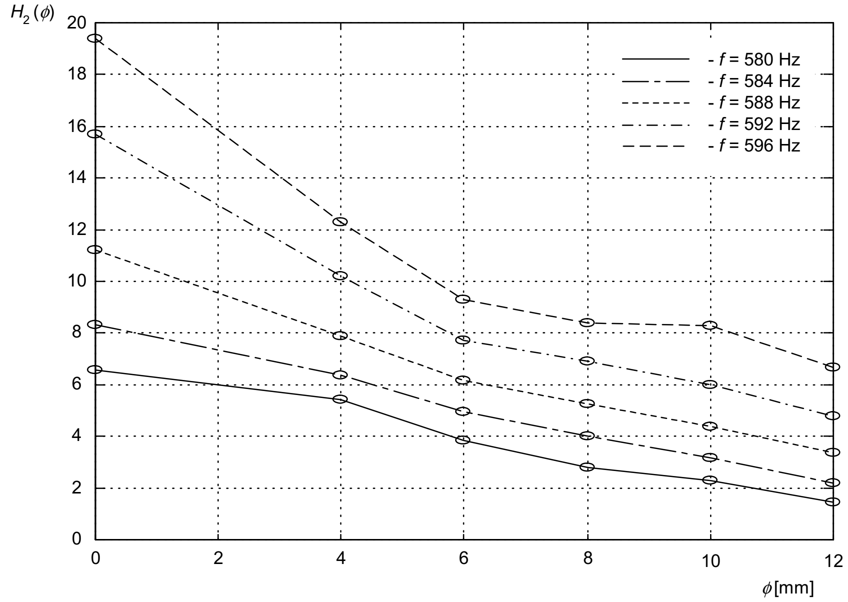


Figure 10. Changing value of amplification factor in function of changes of diameters of the openings in the mast for selected frequencies.

The research has proved that there is a possibility of detecting changes (openings enlarging or/ and changing their form, etc.) in profiles of carbon-epoxide masts. The analysis suggests a possible simple way of defining quality changes. In order to attempt defining the changes in quantitative terms, a broad active diagnostic experiment is necessary, testing a big number of differentiated samples. The frequency range of analysis depends on principal parameters of mast (geometrical dimensions, material characteristics, etc.). The results also show that vibroacoustic methods can be possibly used for evaluating of state of the composite masts in points of fixing construction joints. The main advantage of the applied research method seems to be possibility of submitting a mast to tests without dismantling it, which can significantly lower the diagnosing costs.

6 REFERENCES

[1] Randall R.B. (1987) Frequency Analysis. Brüel&Kjær.

[2] Bendat J.S., Piersol A.G.(1976) Analysis and measurement procedures (in polish). Warsaw, Poland: Państwowe Wydawnictwo Naukowe.

[3] Bendat J.S., Piersol A.G. (1980) Engineering Applications of Correlation and Spectral Analysis. John Wiley & Sons Inc.

[4] Dabrowski Z.(2001), Modelling and Identification of Nonlinear System for Technical Diagnostics. Proceedings of the 8th International Congress on Sound and Vibration, pp. 1083-1090, China, Hong Kong.

[5] Dabrowski Z., Dziurdz J. (2003), Vibroacoustic Analysis of the Changes in Diameter of Holes in Constructional Joints of Composite Masts (in polish). Warsaw, Poland: Mechanical Review, 4, pp. 36-40.

[6] Dabrowski Z., Dziurdz J., Skorski W.W. (2003), Vibration Analysis of The Masts Made of Epoxy-Carbon Composites and Glued Veneer. Proceedings of the 10th Asia-Pacyfic Vibration Conference, Australia, Gold Coast, pp. 75-80.

[7] Dabrowski Z. (2005), Use of Non-linear symptoms in technical Diagnosis. International Journal of Comadem, vol. 8, nr 2, pp. 36-41.

THE INFLUENCE OF CRACK PROPAGATION OF THE BENDED BEAM ON PROPAGATION PERFORMACE OF VIBRATION ENERGY

Grzegorz Klekot and Zbigniew Starczewski

Faculty of Automobiles and Heavy Machinery Engineering, Warsaw University of Technology, Narbutta 84, 02-524 Warszawa, Poland.

Abstract: Possibilities of using the vibroacoustic diagnostic methods for monitoring condition of concrete pre-stressed structures have been researched by the Laboratory of Vibroacoustics, Faculty of Automobiles and Heavy Machinery Engineering. An active experiment was carried out in order to examine the relations between loading (up to damage) of a concrete beam and its vibrations. The research methodology has given an opportunity to assess the possible use of descriptors reflecting propagation of vibration energy for evaluating the stage of failure development or the level of crack propagation. The concept of these descriptors (measures) is based on variable in time energy of processes, proportional to the square of instantaneous value of amplitude. As another assumption of this simple equation is that the defined measures are inversly proportional to the force of impact, it results easy to process the descriptors on the basis of synchronically recorded time signals. The preliminary results have shown that the proposed method can be considered as a supplementary way of detecting damages at an early state of propagation. At the current, preliminary stage of research it can be said that the location of both the impact point and the accelerometer, as well as direction of measured vibrations of the object have proved to be important for diagnostic utility of detectors.

Key Words: Fault monitoring, Vibrations, Concrete structure

1 INTRODUCTION

The materials commonly used in construction are mostly characterized by anisotropy of their strength qualities: unquestionably higher compressive stress is acceptable than tensile, tangential or torsional stress. However, there is a group of several elements submitted to bending, working under a complex stress (floor beams, bridge and overbridge spans, etc.) where load is transmitted mostly by the steel profiles and ferroconcrete structures. Everyday the prestressed concrete becomes more frequently used, means concrete submitted to stressing by flexible distortion of steel bars placed within the concrete mass.

The supporting structures of bridges and overbridges undergo, as any technical objects, exploitation wearing processes. The vehicles circulating in Poland often carry excessive loads which results in speeding up wearing processes of the roads. According to forecasts, in coming years either replacement or major repairs of many bridges will be necessary, as well as building new ones. Thus studies leading to formulation of some efficient methods for evaluating and forecasting technical conditions of these objects are needed.

The Vibroacoustic Laboratory of Institute of Machine Design Fundamentals, Warsaw University of Technology, has inaugurated studies on applying the vibroacoustical diagnosis in evaluation of technical conditions of prestressed concrete structures. Up to date several experiments have been carried out, consisting of controlled overloading some concrete beams working under bending forces until their destruction. The details of the experiments are presented in [1, 2]. The research carried out within the framework of the COST programme are focused on developing vibroacoustic diagnostic methods to be applied to the prestressed concrete structures.

The methodology consists of registering acceleration time courses of impact–forced vibrations of a beam, the beam being both side supported and loaded by the static centrally applied force. This made possible to apply the measures of propagation of vibroacoustic energy developed by the Author and check their potential in evaluating level of damage development. The paper presents some results obtained using the data collected in course of the pilot experiment. The conclusions confirm the direction of further studies to be carried out.

2 THE MEASURE OF EFFICIENCY OF VIBRATION PROPAGATION

In the engineering reasoning real constructions are usually modeled in a simplified way. The dynamic models of one or few degrees of freedom are often considered a satisfactory base for pondering on construction or practical qualities of machines and devices. However, difficulties appear when the non-linear component emerges, which is unfortunately the way technical state of objects changes [3]. Consequently, diagnostic application of the measures based on description of movement of a one degree of freedom system is quite limited. There are different research centres developing methods of forecasting and evaluating the object technical state and trying to make them both efficient and well-anchored in physical bases of analyzed phenomena.

The relationship between the technical state of a mechanic system and the energy dissipated during performance of its functional tasks has been observed [4]. It seems sensible then to search for energetic descriptors to be possibly used in diagnostics of technical objects. So we come to the idea of defining a measure matching a simplicity of expression with an universal quality, and reflecting at the same time the vibroacoustic effect of an input of determined energy [5].

It is commonly known that the average square value of amplitude of a signal variable in time is proportional to the energy of the process characterized by the signal. We can thus examine the following relationship:

$$U \sim \frac{1}{T}\int_0^T \left(x(t)\right)^2 dt \qquad (1)$$

where ***x(t)*** represents the variation course of amplitude of the examined physical quantity in time, while **U** is energy, and ***T*** observation time.

Let us try to use the above statement in our pondering on the vibroacoustic processes. Let us also assume that a description of inputs provoking vibroacoustic activity of the technical object is known in form of ***F(t).*** The measure matching the energy of observed process with the input quantity, and called the measure of propagation of vibroacoustic energy seems to be a valid solution:

$$H = \frac{\int_0^T \left(x(t)\right)^2 dt}{\int_0^T \left(F(t)\right)^2 dt} \qquad (2)$$

If we replace ***x(t)*** by the course of acceleration of vibrations physically registered (measured) for an actual point of the technical object ***a(t)*,** and take ***F(t)*** as the course of input force for the vibrations, we will be able to talk about the measure of efficiency of vibration propagation:

$$H_a = \frac{\int_0^T \left(a(t)\right)^2 dt}{\int_0^T \left(F(t)\right)^2 dt} \qquad (3)$$

The above form is particularly useful while examining the results of experimental studies when the object’s vibrations are forced by impulse. As to damped vibrations, it is enough to take the integration time ***T*** to be longer that the time of vibration disappearing, and the lower integration limit established for the moment of impulse input; subsequently the above quotient will explicitly reflect the relationship between the energy of vibration movement and the particular qualities of an actual object.

One of the characteristics of numeric signal processing is operating on discrete values obtained by discretization and quantization of an analogue signal: so the integrating is replaced by the summation.

$$H_a = \frac{\sum_{i=1}^{N} \left(a(i)\right)^2}{\sum_{i=1}^{N} \left(F(i)\right)^2} \qquad (4)$$

In the above relationship the amplitudes of input force ***F(i)*** and vibration acceleration ***a(i)*** correspond with values of i-th time step. Value of measurement can be calculated on the base of a single impulse or a series of impulses; the only condition is a synchronic processing of both physical quantities used to define the measure.

3 PILOT EXPERIMENTAL STUDIES

The objective of pilot studies was to consider validity of use of the vibroacoustic methods for the purpose of diagnostics of concrete construction elements. A concrete beam supported by two supports (as shown below) was centrically loaded with a force of jerkily increasing values until it broke.

For any of the values of effectuated load, the impulse inputs and the vibration answer of the object were synchronically recorded using six accelerometers located as it is shown in the figure 1 (points 2-7). The vibrations of the beam were activated using an impact hammer equipped with a force transducer; the points of impact are marked 1, while the arrows demonstrate the impact directions. Top values of impulses of input reached 3-5kN.

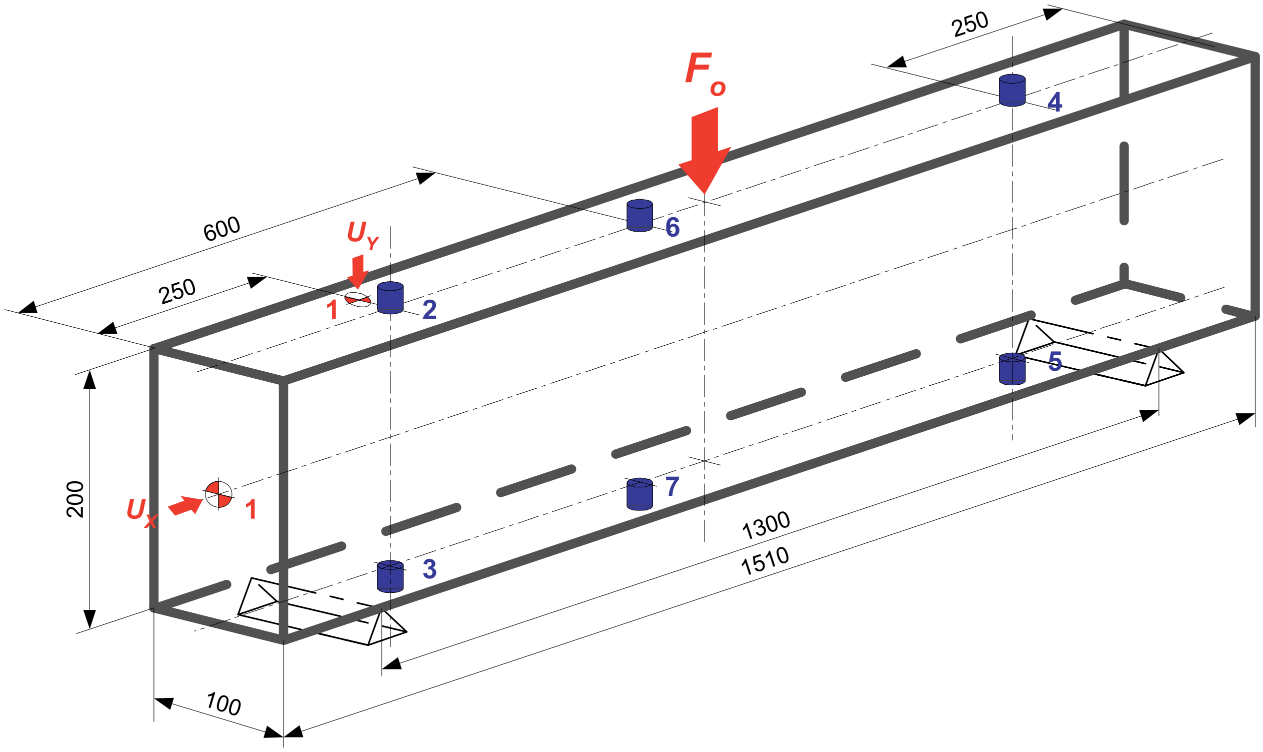


Figure 1. The studied object: location of accelerometers, point of load application and impulse input points.

4 THE RESULTS OF NUMERIC CALCULATIONS

Single numerical values of the measures of efficiency of vibration propagation were calculated according to the relationship (4) as average value coming from a series of 10 subsequent input impulses. The cases of stroke in perpendicular directions were analyzed separately.

The graphs in the figures 2 and 3 show the changes of measures of efficiency of vibration propagation as a function of static load of a beam, obtained by processing the changes of input and the variation courses of acceleration of vibration in fixing points of one of the six accelerators. During the registration of vibration signals of the beam under maximum load, the lead was destroyed, so the graphs illustrating the measure values for vertical strokes have been prepared for static load under 40kN.

The highest numerical values for measures of efficiency of vibration propagation were obtained for the measurement points located near the places of stroke: it seems consistent that structural damping results in lowering of vibration amplitudes in the places situated within a bigger distance from the source. The load input changes to some extent the way the object is supported, and in consequence it explains the maximum without loading and lowering in the function of loading force.

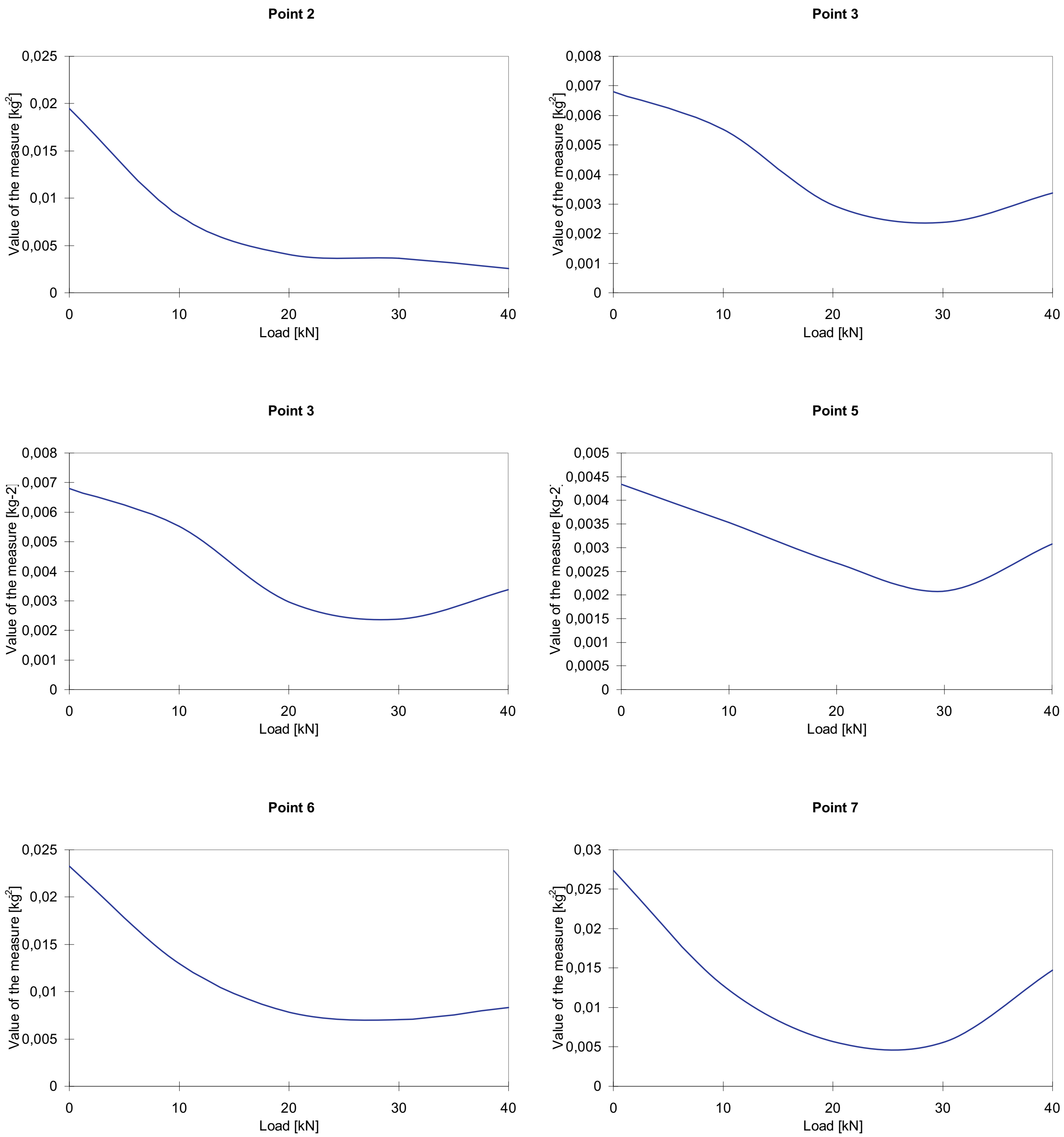


Figure 2. Changes of efficiency of vibration propagation for separate measurement points, vertical input (U_x).

The changes shown in the graphs have a monotonous character only in isolated cases which points to their potential diagnostic utility. On the other hand, there is a lack of an explicit trend in majority of the graphs that suggests a need of deepening the analyze in aspect of this types of application.

Interpreting the graphs and relating them to the object's state become easier if we refer to the information that the first visible cracks (break initiation) were observed under the load of 30kN. In many courses the local extreme corresponding this load can be seen. Quality evaluation makes possible relating the changes of the curve's shape to the initiation of damage.

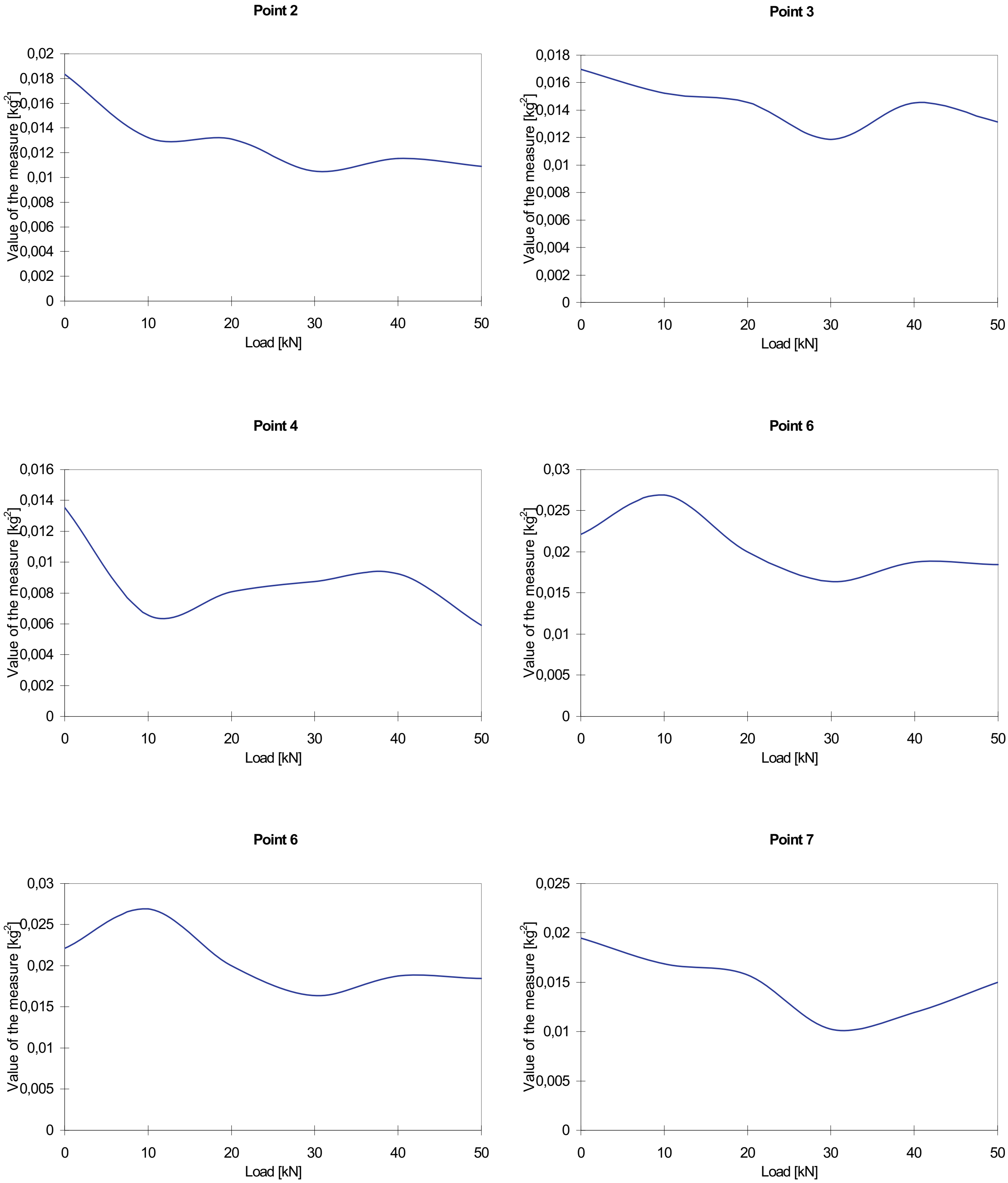


Figure 3. Changes of efficiency of vibration propagation for separate measurement points, horizontal input (U_y).

5 CONCLUSION

The conclusions based on the experiment point to possibility of applying the efficiency measures of vibrations propagation for exploitation diagnostics of concrete structures. However, some additional experimental tests of similar objects seem recommendable, especially of elements made of pre-stressed concrete. Everyday structural beams made of pre-stressed concrete and bended under a load are being more frequently used. Therefore, considering a considerable information value of efficiency measures of vibrations propagations calculated on the base of vibrations induced by horizontally oriented impulses, we would suggest further studies. They would include registration by accelerometers situated at both ends of the beam, processing the motion accordingly with the direction of its induction. Application of bigger number of measurement processors is planned, as well as thickening of changing step of static loading for different qualities of preliminary tension of a stressing element.

All the defects appearing in exploitation process (plastic deformations, cracks, fractures, etc.) change dynamic characteristics of the object, so attempts to apply the already explained measures for evaluating the level of wearing of concrete structures seem justified. Any damage results in modification of the ways of propagation of vibroacoustic energy which necessarily provokes changes of form of the structural vibrations, and consequently, a visible change of numerical value of the measure of propagation of vibrations.

Thanks to virtually free localization of receptors of vibroacoustic signal (accelerometers) and virtually free choice of application point of inducing impulses there should be a possibility of directing the measure towards comparing the properties of actual elements of studied structures, or their fragments. Further selection process can give a chance for crating a tool sensitive to damages of chosen, determinate types.

6 REFERENCES

[1] Radkowski S., Szczurowski K. (2005) The Influence Of Stress In A Reinforced And Pre-Stressed Beam On The Natural Frequency, Workshop of COST on NTD assessment and new systems in prestressed concrete structures, COST Action 534 New materials and systems for prestressed concrete structures. Kielce, 160-170.

[2] Dziurdź J. (2005) Analiza konstrukcji nośnej z materiału niejednorodnego, XXXII Ogólnopolskie Sympozjum Diagnostyka Maszyn, Węgierska Górka, 40, CD.

[3] Dąbrowski Z. (1992) The Evaluation of the Vibroacoustic Activity for the Needs of Constructing and Use of Machines. Machine Dynamics Problems, Vol. 4. Warszawa.

[4] Cempel C. (1989) Diagnostyka wibroakustyczna maszyn, Warszawa PWN.

[5] Klekot G. (2002) Proposed Method for the Evaluation of Sound Insulating Material Efficiency in the Task of Minimizing the Acoustically Onerous Machines, Machine Dynamics Problems, Vol. 26, Warszawa, 5-25.

VIRTUAL TESTING OF A POLYMER SLIDING CONTACT

[1]M. Azeem Ashraf, [2]Bijan Sobhi-Najafabadi, [3]Micheal G. Ellis, [4]Hung-Yao Hsu

[1,2,4]School of Advanced Manufacturing and Mechanical Engineering
Centre for Advance manufacturing research, University of South Australia
Mawson Lakes, SA 5095, Australia
[3]Validation Manager, Schefenacker Visions Systems, Lonsdale, SA
[2]Email:bijan.sobhi-najafabadi@unisa.edu.au Tel : +61 (8) 8302 5018 Fax : +61 (8) 8302 5292

Abstract: The growing competition in the market demands a better performance from any product in terms of availability, reliability and maintainability. The product is expected to perform well beyond its original life expectancy, which requires efficient prediction tools. How long a component can last, requires life prediction methods such as Accelerated Life Testing and *Virtual Life Testing*. *Virtual life testing* method offers a cost effective and fast option to predict the useful life of mechanical components. In this method, it is important to have reliable models to predict different aspects of the component design performance, one of which is wear. Wear is the result of sliding friction between two contact surfaces, which has been the subject of extensive research and investigations. Due to complex nature of the problem, there is no single wear model, which can be adopted for wear prediction. This paper discusses a systematic approach for modelling of dry sliding wear using time domain models. Models are capable of predicting the wear in transient (running-in) and steady state regimes. The predicted wear volume is translated into dimensional changes in the actual geometry to determine the remaining useful life of a contact. Interpolating data obtained from pin-on-disc wear experiments has been used to form the models. The validity of the modelling approach is verified by comparing the predicted results of wear experiments, using the approach described, with that actually measured.

Key Words: modelling, useful life

1 INTRODUCTION

The evolution of market over the past century has shifted the focus of product development efforts from producer to customer. At the beginning of 19th century the market was known to be "producer oriented", namely anything produced in bulk could be sold. Today the customer plays a driving force in design, development and the quality of the product. With frequent advances in technology, customers need products that meet their changing requirements. The competition forces manufacturers to continuously develop new products, which have higher reliability, enhanced useful life and features to replace the ones in market [1]. Moreover, it is expected that products perform satisfactorily well beyond their original life expectancy, which requires tools for predicting the product behavior during their development [2]. A product has to go through a series of development steps namely: Concept Design, System Design, Detailed Design and Testing before it is released to the market [3]. In order to design a reliable product, the last step holds an important position in the product development cycle and often consumes a significant portion of time [4].

Accelerated Life Testing (ALT) is the current prevailing industry practice at the Testing phase of the product development. This is a process, in which prototypes are subjected to accelerated tests specified in the product's specification, so that failure occurs in a short period of time [5]. Based on ALT results, the reliability of the unit is estimated for its normal operating conditions. ALT not only consumes a considerable time of product development cycle, but also incurs a cost through the prototype units utilised for testing.

To grasp competitive edge, the manufacturers need to limit the amount of physical testing and design iterations thus minimizing the time-to-market incurred ALT. One such direction is "Virtual Life Testing" method for product reliability evaluation. The method uses mathematical models of the product instead of using the physical prototypes. This method offers a cost effective and fast option to predict the life expectancy of mechanical components using powerful computers and software [6]. Sliding wear is one of the important physical characteristics of , which can be modeled and used in VLT.

Sliding wear can be defined as the progressive material loss, which occurs on the contact surfaces of components that experience a relative sliding motion [7]. Wear of mechanical components is one of the major failure modes encountered in sliding contacts. It is a critical factor influencing the service life span of components. The ability to predict the wear and hence the useful life of sliding members is important in redesigning or proposing a replacement schedule for faulty components.

As illustrated in Fig. 1, mechanical wear can be classified in four major modes; Adhesive, Abrasive, Erosive and Fatigue. These wear modes are not mutually exclusive and in most cases they can be transformed into each other depending on the operating conditions. Adhesive wear is the most commonly occurring wear [8].

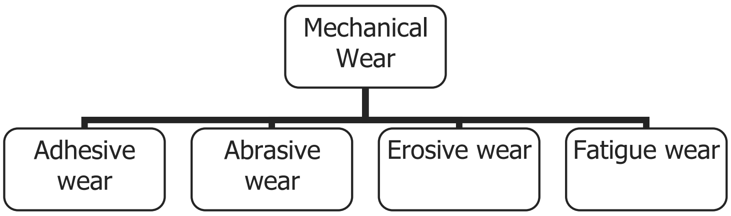


Figure 1: Major wear modes

2 MODELLING OF SLIDING WEAR

There are over 300 wear models documented in literature [9]. The dependence of sliding wear on dynamic factors like; load, speed, temperature, time, contact geometry, surface roughness, oxygen availability, lubricant chemistry, and material surface compositions makes it difficult to have a single comprehensive representation [10]. In spite of the existence of such a large number of wear models, there is still no single wear model that can be used for all situations [11].

In general wear volume loss can be defined as

$$V = f(P, S, \mu, H, t, R) \tag{1}$$

where P, S, μ, H, t, and R represent the applied normal load, relative sliding speed, coefficient of friction, hardness of the softer material, sliding time, and the size of contact respectively. Normally, for wear modeling it is assumed that the wearing action will only occur on the softer of the two sliding surfaces. The two wear models used in this research are Archard's wear model and Yang's wear model, which are described next.

2.1 Archard's wear model

The wear model proposed by Archard has been extensively used for wear coefficient calculations . The wear volume loss, V, for a sliding contact is given as

$$V = K_s \frac{PL}{3H} \tag{2}$$

L and P are sliding distance and normal load applied to wear surfaces respectively. H is the Brinell hardness of the softer wearing material and K_s represents the steady-state wear coefficient. Therefore, for known values of V, P, H and L the standard wear coefficient can be calculated [12].

2.2 Yang's wear model

Most of models treat the wear in steady-state and not considering the transient (running-in) wear. A wear model was recently presented by Yang, which takes into account the initial running-in wear of a contact [13]. The transient wear volume can be calculated from.

$$V_t = A(1 - \exp^{-BL_t}) \tag{3}$$

V_t and L_t represent the transient wear volume loss and the sliding distance accordingly. A and B are constants, whose values are determined by fitting a curve using equation 3 through the experimental values of wear volume loss and sliding distance. Subsequently, making use of equations 2 and 3, the wear coefficient can be computed using equation 4

$$K_s = \frac{3HA}{PL}[1 - \exp^{-BL_t}] \tag{4}$$

The wear coefficient thus calculated will follow an exponential curve, which indicates that the virgin contact surfaces have higher number of asperities wearing as a result of running-in wear. After the running-in stage the wear coefficient follows a steady pattern for the given conditions. Once in the steady state wear region, equation 2 can be used to model the wear [13].

3 LIFE PREDICTION METHDOLOGY FOR SLIDING CONTACT

Traditionally, the conservative safe life of critical mechanical components for (time or usage) was estimated statistically. Although, this method still works well, however the current evolution of diagnostic and prognostic techniques, coupled with the development of advanced computing systems can provide accurate design and prediction tools. Model based prognostic methods have proven to be useful for life expectancy prediction of mechanical components [14]. The life prediction model for the sliding polymer contact in consideration goes through the steps, which are described in the following sub-sections.

3.1 Experimental setup and procedure

Pin-on-disc method is commonly used for determination of the wear coefficient of a sliding contact. The pin is usually the wearing member and has a lower hardness value. However, instead of standard pin-on-disc in the present case the experimental contact geometry was designed to mimic the geometry of the real contact as it appears in the final product. Figure 2 (a) depicts the actual geometry consisting of three sliding contact pairs.

The test setup is demonstrated in Figure 2 (b). Referring to Figure 2 (b) 'Track' is used as the disc and the 'Slider' correspond to pin when it is compared to conventional pin-on-disc method Slider' is mounted on a rotary pneumatic cylinder, which can be revolved in clock and anti-clock wise directions relative to the fixed Track. It consists of three contact pairs. There is a 430 N dead load applied on the Slider as the normal load, which is assumed to be equally distributed across each of the three contact pairs (~143 N). The angular stroke of the Slider is 32.3° as defined by the operating range of the mechanism. Thus the linear sliding distance can be computed for wear volume loss calculations.

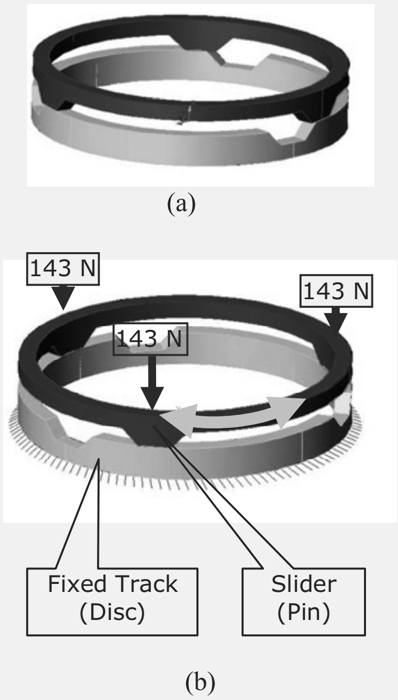


Figure 2 (a) Configuration of actual contact (b) Experiment setup

Slider is made of a 33% glass reinforced, heat stabilised black nylon copolymer resin density 1400 kgm^{-3} and tensile strength 172 MPa) commercially marketed by DuPont as Zytel 72G33HS1L and the sliding Track is made of BK15945% glass reinforced modified polyethylene terephthalate (density 1700 kgm^{-3}, tensile strength 186 MPa and hardness $95HR_M$ ($120HR_R$)) commercially marketed by DuPont as Rynite 545 NC010.

Track is made of relatively harder material than the slider therefore it is assumed that only the slider will be worn in wearing process. Wear volume loss is measured at sliding distances of 25, 50, 75, 100, 125, 225, 325, 425, 525 and 625 metres.

The service life of the mechanism is 35000 cycles comprising folding 32.3° and reverse fold to the original position. The sliding distance of each contact of Slider is 30mm through the fold – reverse fold cycle. The height of Slider contacts has direct influence on the service life of the mechanism described above. Therefore, the purpose of virtual testing is to model and simulate service life of one of the three identical Slider contacts. Since the Slider contact pairs are identical, modeling one is sufficient. In order to have sufficient data points throughout the expected service life of the mechanism it was decided to obtain the data points in 25 m increments (787 folding cycles). When 3935 folding cycles or 125 m linear sliding distance was reached, the measurement frequency was altered to 100 m increments, which is equivalent to 3148 folding cycles.

The sliding speed during testing was set at 6.3 x 10^{-3} ms^{-1} (12.3 $°s^{-1}$). The normal load of 450 N applied as per the mechanism design condition. The ambient temperature during testing was controlled at 22 ± 1 °C.

Figure 3 depicts the general approach that has been adopted for prediction of height decay of the sliding contact pairs [15].

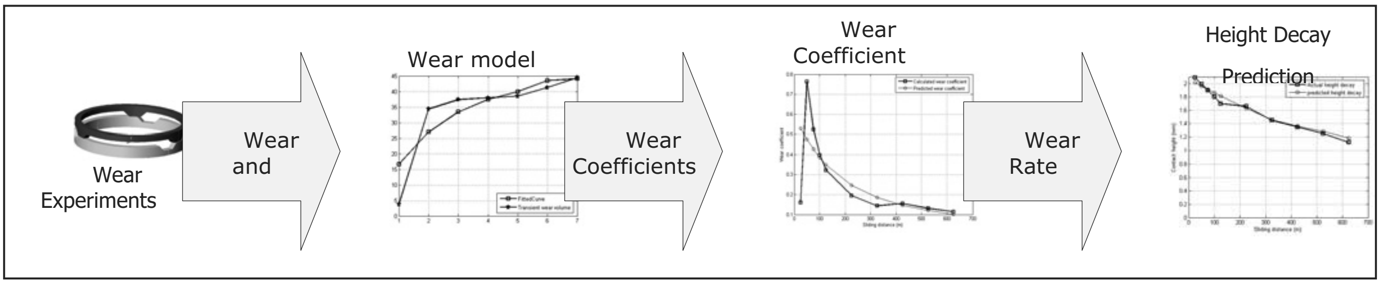
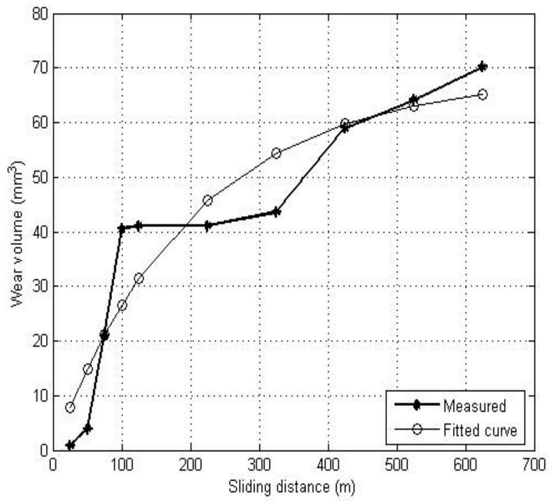


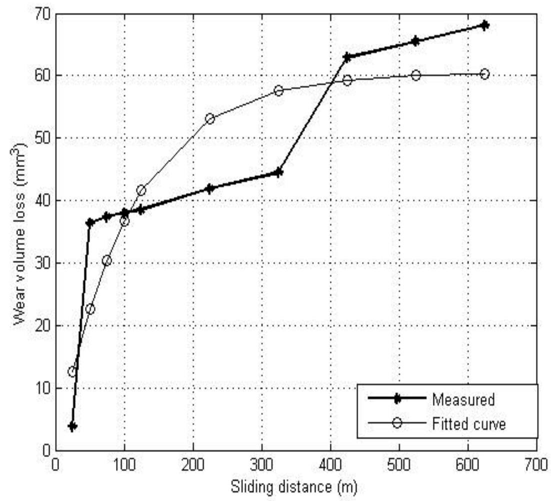
Figure 3: Modular virtual testing approach for dry sliding wear in time domain time

3.2 Calculations of Yang's constants

In order to obtain constants A and B, which are required for calculation of the wear volume loss in Equation 3, two experiments were carried out with two sets of slider and track. Curves were fitted through each set of experimental data. The experimental results of wear volume loss versus sliding distance together with the fitted curve for each of the two experiments are illustrated in Figure 4 (a) and (b).



(a) Experiment 1



(b) Experiment 2

Figure 4: Wear volume versus the sliding distance and the curve fitted around the data

The fitted curves in Figures 4 (a) and (b) show the transient (running-in) wear reaching a steady-state. From each plot it can be observed that the running-in wear extends from sliding distance of 0 to approximately 400 meters. Beyond this distance, the wear tends to reach steady state condition. Due to the limitations of the experimental setup, further verification of steady state regime could not be carried out. The spikes of the experimental curves, above or below the fitted curves may be attributed to the small step size of sliding distance for which the wear volume loss was recorded. Such spikes would be masked out if larger increments were considered. The experiments are carried out over varying sliding distances at a constant speed, thus the wear volume loss results are in time domain.

Table 1

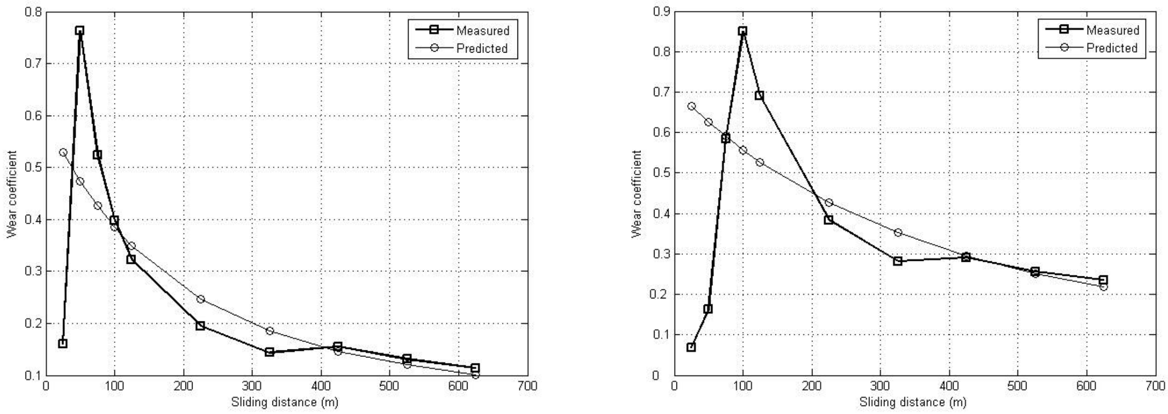
Constants A & B for transient wear of the sliding contact

	A	B
Experiment 1	60.4482	0.0093
Experiment 2	68.1899	0.0049

The values of calculated constants A and B for the two experiments are given in Table 1, which are quite high compared to constants calculated for the model presented by [13] for aluminum composites. This is in agreement with the difference in material properties of metal and polymers. A glass re-enforced polymer composite is much more prone to wear as compared to metals.

3.3 Wear coefficient calculations

Both measured wear volume loss and fitted curve data from previous section were used to calculate the wear coefficient values using equation 4. The graphs of both measured and predicted wear coefficient versus sliding distance for each of the experiments explained in the previous section are shown Figure 5 (a) and (b). Both of the wear coefficient plots resulting from the two experiments indicate a very high initial spike from a low to very high wear coefficient. As can be seen from the Figures 5 (a) and (b), the initial distance increments results in a very small wear coefficient. The reason being, once the initial asperities have released abrasive particles in the contact region, the wear coefficient increases to a high value, which later slowly decays to a low but steady wear coefficient.

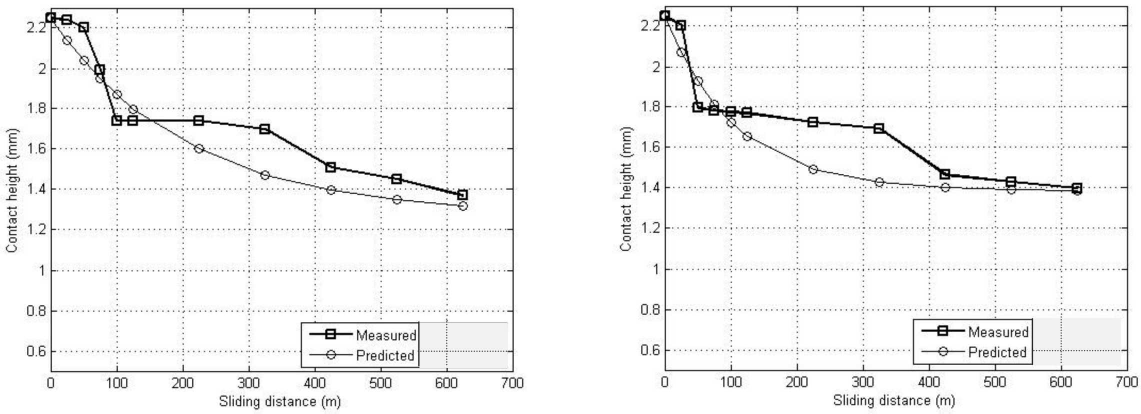


(a) Experiment 1 **(b) Experiment 2**

Figure 5: Experimental and predicted wear coefficients versus sliding distance

.

3.4 Contact height decay



(a) Experiment 1 (b) Experiment 2

Figure 6: Actual and predicted height decay vs sliding distance

Given the geometry of the wearing member, the wear volume loss at each distance increment can be translated into the height loss of the contact. This in turn gives the contact height decay as a function of sliding distance or number of cycles. The decay rate can be compared with performance specification of the product. This information helps designer or maintenance engineer to decide on redesigning or scheduling replacement of the worn component.

Figure 6 (a) and (b) illustrate the measured and predicted height decay of the wearing member versus sliding distance for the two experiments. The measured and predicted curves correlate well, showing the accuracy of modeling approach. The graphs demonstrate rapid height decay at the initial distance increments that is in conformance with the initially high wear coefficient as discussed in section 3.3. The subsequent distance increments demonstrate a gradual decrease in the wear rate till it reaches a steady state condition. The predicted decay curves - are based on the predicted wear volume data from the calculations in section 3.2, showing no decay as steady state is reached. In contrary, the decay should continue beyond the steady state but with a different governing model, which could not be demonstrated due to limitation of the experimental setup as was explained before.

4 CONCLUSIONS

Under dry sliding conditions, the ability of wear models to predict wear conforming to experimental values varies greatly for different polymers and test conditions [14].

During the experiments, it was observed that the actual load on the slider was not uniformly distributed between the three contact pairs, however it was consistent between the two experiments. The varying load between the contacts caused them to wear unevenly, affecting the overall geometry of the Slider. Therefore, it can be the deduced that the individual wear of a contact is influenced by the wear of other contacts. This may prove to be a challenge to understand and simulate this problem in future.

The design of Slider's contact causes the leading edge to wear at a faster rate than the trailing edge. Determining the relationship between geometry and material properties for wear modeling is important, which will be studied in future.. Metrology techniques have been marginal during these experiments due irregular wear caused by mechanism geometry. A standard pin-on-disc configuration offers more options with respect to material loss measurements.

Slider's contact geometry used in these experiments was limited by height to keep it in accordance with the actual design of the end product. Therefore, the height was not enough to carry on the experiment beyond the running-in phase of the model. Moreover, the environmental and loading conditions were set only for this particular product design. Further experiments will be carried out with different geometries, loading conditions and environment variables to establish a more generalized wear model for these materials.

The models and predictive methodology presented by Yang [13, 16] were based on confirmation experiments carried out for aluminum composites only. The sliding distances selected and the corresponding wears found in the current experiments have no comparison with the values in [13] due the difference in the nature of materials. Since, limited literature on polymer wear was found by the authors, therefore it is difficult to benchmark the findings.

The virtual life testing approach depicted in Figure 3, is not completely virtual and there is some experimentation involved in establishing the models. In order to standardize designs and manufacturing equipment to optimize the operating cost, industries tend to design components that could fit in a wide range of products with minimal variation. Once successful wear models have been established corresponding to a particular product and operating conditions, they may be adapted to a range of similar products.

5 REFERENCES

[1] Schuman, C. A. and Brent, A. C. (2005), Asset life cycle management: towards improving physical asset performance in the process industry. *International Journal of Operations & Production Management*, **25**(6): p. 566-79.

[2] Impelluso, A. (2002) A simplified approach to capital asset life extension. *Power Engineering (Barrington, Illinois)*, **106**(6): p. 45-46.

[3] Product Design and Development Chapter 3 Karl T. Ulrich and Steven D. Eppinger2nd Edition, Irwin McGraw-Hill, 2000.

[4] Thomke, S. and Bell, D. E., Sequential Testing in Product Development, *Harvard University, Graduate School of Business Administration*, Boston, Massachusetts 02163.

[5] Lydersen, S. and Rausand, M. (1987), A systematic approach to accelerated life testing, *Reliability Engineering*, **18**, 4: p 285-293.

[6] Hegadekatte, V., Huber, N. and Kraft, O. (2005), Finite element based simulation of dry sliding wear. *Modelling and Simulation in Materials Science and Engineering*, **13**(1), p. 57-75.

[7] Williams, J.A. (1999), Wear modeling: analytical, computational and mapping: a continuum mechanics approach. *Wear*, **225-229**, p 1-17.

[8] Rabinowicz E. (1995) Friction and wear of materials, New York, John Wiley & Sons.

[9] Meng, H.C. and Ludema, K.C. (1995), Wear models and predictive equations: their form and content. *Wear*, **181-183**(2), p. 443-57.

[10] Hutchings I. M. (1992), Tribology: Friction and Wear of Engineering Material, London, Edward Arnold Publishers.

[11] Williams, J.A. (2005), Wear and wear particles - Some fundamentals. *Tribology International*, **38**(10): p. 863-870.

[12] Bhushan B. (2001) Modern Tribology Handbook, Boca Raton, CRC Press.

[13] Yang, L.J. (2005), A methodology for the prediction of standard steady-state wear coefficient in an aluminum-based matrix composite reinforced with alumina particles. *Journal of Materials Processing Technology*, **162-163**: p. 139-148.

[14] Watson, M., Byington, C., Edwards, D. and Amin S. (2005), Dynamic Modeling and Wear-Based Remaining Useful Life Prediction of High Power Clutch Systems. *Tribology Transactions,* **48**, p 208-217.

[15] Franklin, S.E. (2001), Wear experiments with selected engineering polymers and polymer composites under dry reciprocating sliding conditions, *Wear*, **251,** p. 1591–1598.

[16] Yang, L.J. (2005), A test methodology for the determination of wear coefficient, *Wear*, **259**, 7-12,: p. 1453-1461

Acknowledgement

The author would like to thank the support of this project by Schefenacker Vision Systems Australia (SVSA) for providing the material and experiment facilities.

UNDERSTANDING FAILURE RISK IN ELECTRICAL MACHINES

MIKE DAVIS

Machinemonitor Pty Ltd

Abstract: Most operators of business critical electrical machines experience unexpected electrical failure events from time to time. The consequence of these events is typically significant losses resulting from down-time and recovery costs. This paper outlines an approach to establish electrical failure risk in a given machine through understanding machine condition in the context of design characteristics, maintenance and operational histories. Over the last decade extensive data collecting has been undertaken on upwards of 500 high voltage machines to establish risk profiles and a knowledge database. The information gained from the machine modelling has allowed users to make more informative judgments on machine condition, risk and maintenance requirements. Two case studies will be presented one from a Hydro Generating Plant and another from a Minerals Processing Plant to demonstrate how business risk cost can be reduced through modelling and identifying failure risk in electrical machines and targeting maintenance toward mitigating the risk.

Key Words: Business Critical, Down-time, Risk Profile, Hydro Generating, Modelling, Maintenance.

1 INTRODUCTION

A project was developed in 2001 to establish a web based management portal for electrical machines which would permit machine owners to enter machine design operating and condition monitoring data for the purpose of modeling electrical machine failure risk. A common standard methodology was established to asses and analyse machine data for the purpose of establishing and displaying a failure mechanism matrix for individual machines. Once it was understood which failure mechanisms were predominant in a given machine it then became possible to rank machines based on the extent and severity of the identified failure mechanism and what maintenance action would be required to mitigate the failure mechanism(s) and hence risk of failure.

2 THE “RISK MODELING” METHODOLOGY

The six-step high voltage electrical machine risk assessment tool is summarised below. The functionality of this assessment tool takes the machine maintainer through the following simple steps to establish risk profiles on stator and rotor windings for individual high voltage machines. The risk profiling is an on-going life of machine process and will afford the machine owner the opportunity to economically target maintenance to the areas of greatest risk in the monitored machine population.

2.1 Machine Identification

Establish a machine database detailing all technical information on the particular machine. Information collated included;

- Full machine name.
- Manufacturer.
- Machine type i.e. Induction Motor, Synchronous Motor, Steam Turbine Generator or Gas Turbine Generator.
- Speed.
- Voltage
- Output KW or KVA.
- Year of Manufacture.
- Stator Insulation Class, B, F or H.

2.2 Maintenance History

Collate the maintenance history on the machine. This step lists all maintenance activities on the machine with particular focus on major maintenance that may have changed the material content in the machine.

2.3 Operational History

Collate operational data. This following information is required;

- Machine state, In-service, off-line or disassembled.
- Number of starts.
- Operating hours.
- Operating environment i.e. Wet/Contaminated or abrasive environment.
- Events i.e. asynchronous operation, over excitation, stalls, locked rotor etc.

2.4 Design Data

Collate design data. In this collation a thorough review of design data is undertaken with attention being paid to flux and current densities, ground wall stress levels, mechanical assembly and voltage stress distribution. These parameters are then compared against normal engineering design limits. The following data is required,

- Stator insulation – Asphaltic Mica, Mica folium, Polyester/mica splittings, Polyester/mica paper or Epoxy mica paper.
- Winding data including coils, groups, turns and cross-sectional area & stator core dimensions
- Stator winding connection Y orΔ.
- Number of parallels circuits.
- Semi-conductive coating YES/NO.
- Global VPI of stator YES/NO.
- Stator winding type.
- Temperature monitoring YES/NO. Nominal hotspot temperature (stator).

2.5 Data Analysis

Integrate data into machine database. The data collated in steps 2, 3 and 4 is integrated into a machine database to establish the likely predominant failure mechanism(s) that could exist in the machine. These failure mechanisms are presented in descending order of likelihood.

2.6 Verification Testing

Based on the likelihood of a failure mechanism a testing regime is put into place to collect objective data as to the presence and extent of the identified predominant failure mechanism(s). Testing can be selected on the following criteria;

- Usefulness - which test is going to give the most meaningful data to assist in making the risk judgment?
- Risk - which test has the lowest risk for damage to machine?
- Time - which test is the most time efficient test?
- Machine State - which test is best considering the state of the machine?
- Machine History – considering machine history which test is best?

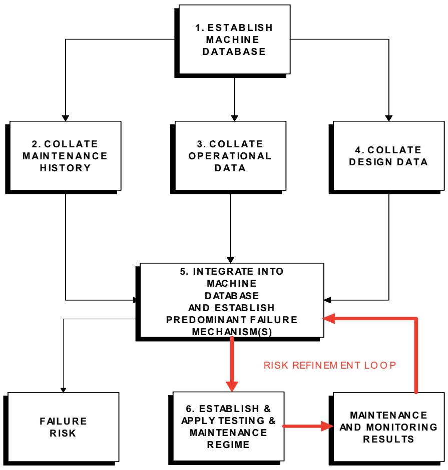


Figure 1 Risk Assessment Process

3 FAILURE MECHANISM AND RISK RANKING

The analysing of machine and test data permits the rank the failure mechanism(s) in the machine stator winding and also ranks overall risk of failure for the machine stator and rotor windings. The failure mechanisms are ranked on predominance through a confidence factor from 0 to 1, 0 meaning there is no evidence of the presence of failure mechanism and 1 meaning the mechanism is definitely present. The process output is illustrated in figure 2.

Risk ranking required development of scores based on severity of the modeled failure mechanism matrix and consideration of the operating environment and design stress raisers in the machine. It became apparent from the entire list of possible electrical failure mechanisms in a machine some develop rapidly whilst others develop slowly in some examples over decades. Elevated risk is typically associated with the presence rapidly developing or advanced failure mechanisms. Risk rankings are expressed in percentage terms and further divided into quartiles.

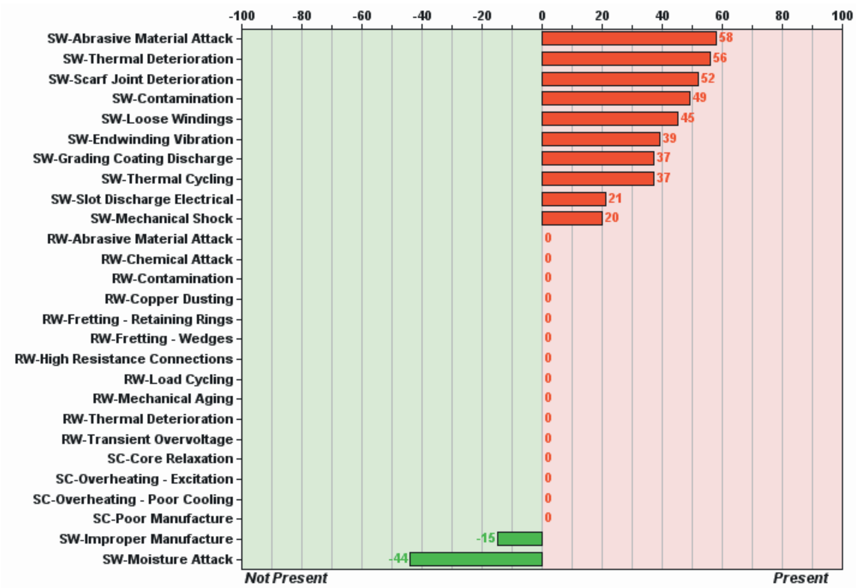


Figure 2 Failure Mechanism Matrix from 50 year old Steam Turbine Generator

4 RISK IMPLICATIONS

Having ranked machines according to risk the next step was to relate a risk ranking to a maintenance action. The relationship between risk ranking and maintenance requirements is tabulated below

Table 1

Risk Implication – Stator Windings

Percentage	Risk Quartile	Risk Implications
1 -50%	1st & 2nd	Associated with reliable machine condition no remedial action required
51-75%	3rd	Low level failure mechanisms present. Plan medium term strategic maintenance
70-90%	4th Level 1	Predominant failure mechanism(s) present. Actively monitor and trend in accordance with specific Inspection Test Plan (ITP)
91-98%	4th Level 2	Susceptibility to unexpected failure. Avoid operational excursions. Plan inspections and maintenance. Check availability of spares.
98% +	4th Level 3	Immediate Maintenance Required

5 ESTABLISHING RISK COST OF ELECTRICAL MACHINES

Operating electrical machines in business critical applications exposes the business to losses associated with unexpected failure. These losses include opportunity cost due to loss of production and equipment recovery cost. Annualised risk cost can be calculated by;

Risk Cost = (η x b) c

Where η = annual probability of failure

b = cost of failure (including downtime and reclamation costs)

c = risk %

6 CASE STUDY ONE – REDUCING BUSINESS RISK COST

A 3.2MW ,11kV, 6 pole wound rotor induction motor operated on milling duty at "the heart" of this cement manufacturing plant for 10 years. No spare was available and down-time cost amounted to $40k per hour. Baseline inspections and risk modeling indicated a 4th quartile level 3 risk at 98% contributed to by loose windings in the stator slots. Modern high voltage machines have a low tolerance to this kind of failure mechanism due to the non flexible nature of epoxy based insulation used in the stator insulation system.

- The risk cost associated with the above condition is expressed as;
- Risk Cost = (η x b) c
- Risk Cost = 0.5(400000.24.60) 0.98
- Risk Cost = $10.944M
- Where
    - η = 0.5 (annual probability of failure on measured condition)
    - b = cost of failure (including downtime and reclamation costs)
    - c = risk %

Maintenance was scheduled on the machine which including removing the machine, transporting to a repair facility and re-wedging the stator. After this maintenance a revised risk model was developed reflecting the maintenance and confirmation testing (Partial Discharge). The revised risk was modeled as Quartile 2 at 42%

- The risk cost associated with the condition after maintenance is expressed as;
- Risk Cost = (η x b) c
- Risk Cost = 0.2(400000.24.60) 0.42
- Risk Cost = $1.209M
- Where
    - η = 0.2 (annual probability of failure on measured condition)
    - b = cost of failure (including downtime and reclamation costs)
    - c = 0.42 %

This experience demonstrated the effect of reduction of business risk cost by simple targeted maintenance. The total reduction in business risk cost for a $100k maintenance investment was in the order of $9.7M.

7 CASE STUDY TWO – TARGETING MAINTENANCE IN A POPULATION OF AGED MACHINES

This hydro generating business operates some seventy generators aged up to 70 years old. In an environment of diminishing maintenance resources and exposure to increasing failure rates there was a requirement to target maintenance toward generators which require attention. The risk modeling process was applied to the entire generator fleet to benchmark machines on risk. Extensive data mining was undertaking in the business to understand the design, maintenance and operating data for each generator. The ranking summary for the machine population is represented below;

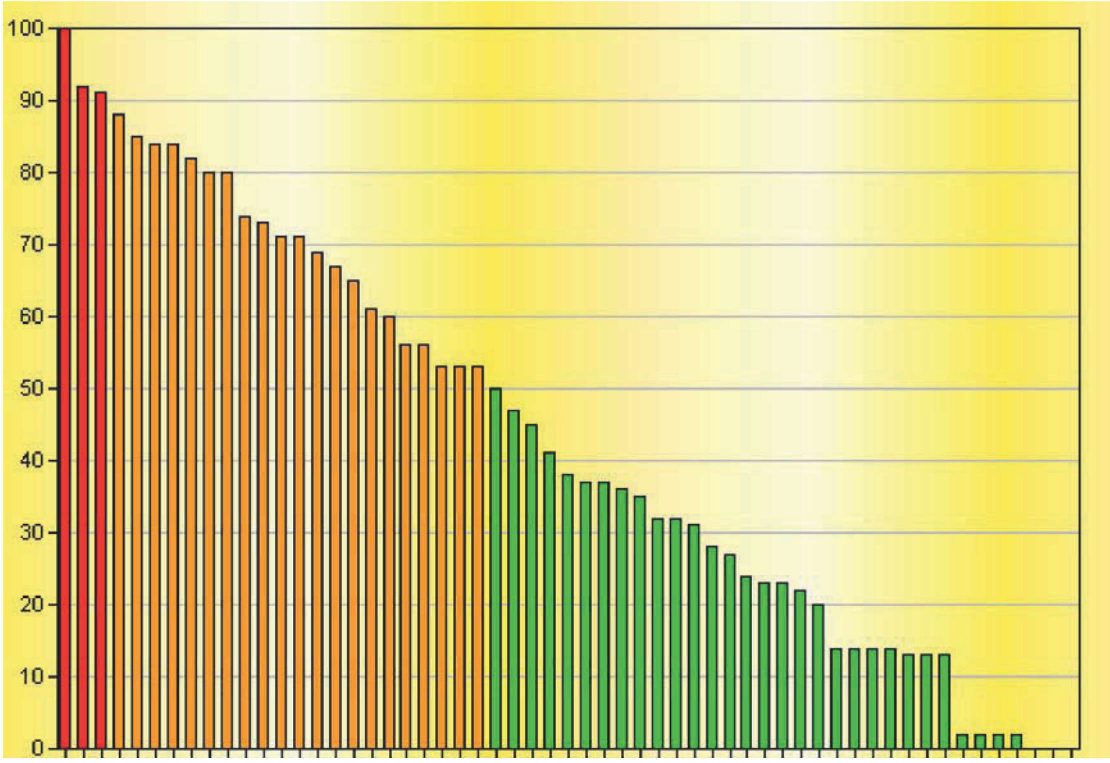


Figure 3 Failure Mechanism Matrix from 50 year old Steam Turbine Generator

As a result of the process one generator was identified as requiring immediate maintenance with an additional two machines requiring maintenance in the short term. In the order of 75% of the machine population returned risk rankings associated with reliable operation.

8 CONCLUSION

Based on the experience of the last five years it has been demonstrated that using common standard approaches for risk modeling across a machine population permits machine owners and maintainers to rank machines based on risk ratings. Furthermore once machines are ranked and the problems driving risk understood significant maintenance and risk cost savings can be achieved through mitigating risk by applying targeted maintenance toward that low percentage of machines with elevated failure risk.

AD-HOC TO BEST-PRACTICE – THE ROADMAP TO ACHIEVING BEST-PRACTICE MANAGEMENT OF CONDITION MONITORING DATA

Adam Van Dyck [a]

[a] *Product Development Centre, CCI Pope Pty Ltd, 29 Rosegum Close Warabrook, NSW, 2304, Australia.*

Abstract: It is widely accepted that predictive maintenance is an essential component of an effective maintenance program. As organisations increase the quantity of predictive maintenance, it becomes increasingly difficult to manage the huge volumes of data generated through the various condition monitoring techniques from potentially many different sources.

There are many shortcomings with the status quo, whereby condition monitoring reports are organised using a manual or electronic filing system. Many organisations struggle to make this information easily available across their organisation. Reports can be difficult to find and searching through these reports is often a time consuming manual process. Collecting the data from various condition monitoring techniques for a particular asset can involve manually searching through many filing systems. Seeing the 'bigger picture' of overall plant health is virtually impossible. While modern Computerised Maintenance Management Systems (CMMS) are more than capable of managing the maintenance program, they are not ideally suited to managing high volumes of condition monitoring data. The absence of a suitable information system is often a barrier in the transition from status quo to best-practice.

Short term benefits are immediately realised when organisations adopt a best-practice approach to condition monitoring data management. An instantly accessible and exhaustive condition monitoring history for each asset is available, enabling the most informed maintenance decisions to be made. Potential problem areas are communicated to the relevant personnel immediately and overall plant health can be easily determined at any point in time. Over the longer term, these organisations are able to extract an extra layer of value from their condition monitoring data management systems. This may include benchmarking the health of similar assets across the organisation as well as identifying both effective and ineffective maintenance practices.

In this paper the author outlines the features and benefits that a best-practice condition monitoring data management system should provide. The people, processes and information systems required for best-practice are discussed in detail. Finally, the author suggests a roadmap that can be followed by those organisations wishing to achieve best-practice data management. Potential challenges that may arise are identified to allow risks to be managed through the implementation process.

Key Words: Predictive Maintenance, Data, Information, System, best-practice, condition monitoring

1 INTRODUCTION

Developing systems and processes to manage condition monitoring data is a significant issue in the overall maintenance program. However, while resources are readily invested in areas such as commissioning CMMS systems and developing maintenance strategies, the issue of condition monitoring data is often left unaddressed. By default the condition monitoring data that flows into the organisation is unmanaged; resulting in significant hidden costs due to time wasted searching for this data and the inability to leverage its full potential.

For the organisations that make the commitment to develop robust condition monitoring data management processes there is a significant prize in terms of time and money saved and increased plant reliability.

2 THE STATUS QUO

2.1 Silos of Data

In many organisations, Condition Management data management is an ad-hoc process. Typically, condition monitoring reports flow into the organisation from various sources, including internal condition monitoring groups, third party contractors, OEM's and on-line systems. Once delivered, these reports become the responsibility of various individuals or departments within the organisation. There is a reliance on the diligence of these individuals to file reports appropriately so that a permanent record is kept and that it is accessible to others within the organisation. This may be done by manually filing a paper report in a filing cabinet or saving a softcopy to a known location on a central fileserver. The end result is often a set of disconnected 'silos' of condition monitoring data across the organisation. Once a report is filed this information is essentially lost to the remainder of the enterprise. It becomes significantly more difficult and time consuming to conduct any higher level enterprise wide analysis of that data. When individuals change jobs or leave the organisation, the manual report filing processes fall over, resulting in further effort to re-establish these processes and the likelihood of an incomplete condition monitoring history for the organisation's assets.

2.2 Inspection-Centric Condition Monitoring Reports

Condition monitoring information is typically delivered to maintenance personnel in an "Inspection Centric" format. That is, it is usually a written report in Microsoft Word, Excel or PDF format that is specific to a single inspection technique, for example, Thermography. A well written condition monitoring report will summarize the problem areas of plant and outline the recommended corrective / preventative actions at the top of the report. This allows the maintenance personnel to flick through the first few pages of the report and action any necessary preventative or corrective maintenance recommendations before filing it. However, often the recommendations in the report are based only on the results of that specific inspection using a particular inspection technique. They may not take into account the holistic health of the component by considering other predictive maintenance data that is available. To do this requires a "Component-Centric" view of the condition monitoring data. That is, For a specific component, maintenance personnel need to see an entire set of current and historical condition monitoring data across all condition monitoring techniques. If the condition monitoring data is delivered and stored in an "Inspection-centric" format such as non-divisible written reports, then maintenance personnel will need to manually find and collate information from appropriate reports, which is a time consuming and potentially error prone process.

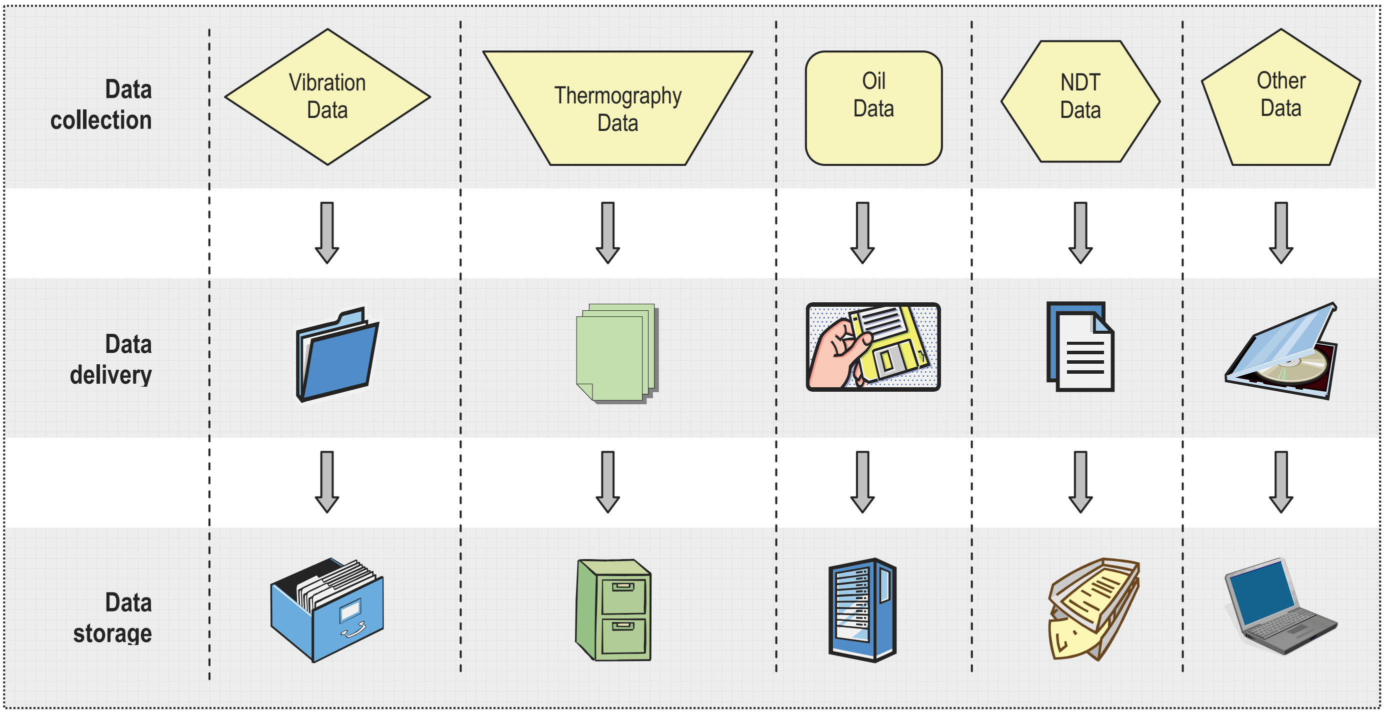


Figure 1: Typical ad-hoc condition monitoring data management results in 'silos' of condition monitoring data. It is difficult to share or collate condition monitoring information across the organisation. Obtaining a holistic view of the current health of all assets can be almost impossible.

2.3 The Limitations of the Status Quo

The ad-hoc storage of "Inspection-centric" condition monitoring reports inhibits an organisation's ability to quickly access and further leverage the condition monitoring data.

Inability to quickly understand current plant condition and identify trouble spots: It becomes difficult for organisations to quickly and easily obtain an overview of the current health of the assets in an area of plant. For example, to understand which assets in the plant are currently in poor or critical condition, a maintenance manager would need to manually collate and analyse all of the relevant recent condition monitoring reports.

Inability to quickly retrieve an exhaustive history of condition monitoring data and trends for a particular component: Often, maintenance personnel will need to review an exhaustive history of the condition monitoring data for a specific component. Take for example a vibration monitoring report that indicates that a particular gearbox may be approaching failure. A decision needs to be made to either allow the gearbox to run or to schedule preventative maintenance. To make the most informed decision the maintenance personnel should review all relevant condition monitoring data for that component. This may include reviewing historical vibration monitoring reports for the suspect gearbox to see if this was the first reported indication of the potential failure or whether the failure indications have been trending upwards over some time. Additionally, relevant oil analysis, thermography and NDT data for the suspect gearbox should also be reviewed for any supporting or contradicting evidence of an impending failure. For organisations that take an ad-hoc approach to storing condition monitoring data, the maintenance personnel will be required to manually find the relevant reports and then collate and analyse this information. In many cases, this work is too difficult and time consuming so the maintenance decision is made based purely on the evidence presented in the most recent condition monitoring report.

Inability to add value to condition monitoring data: Organisations lose the ability to turn condition monitoring data into knowledge through data mining. For example, an emerging trend of poor health of similar assets across the enterprise may identify the need for a maintenance strategy review or even a design review. An ad-hoc approach makes it infeasible to trowel through condition monitoring data across the entire enterprise and to extract the knowledge contained within the data.

3 WHAT IS BEST-PRACTICE CONDITION MONITORING DATA MANAGEMENT?

Best-practice condition monitoring data management is a system of three equally important and inter-dependent components working in synergy.

- The underlying ***information system*** is the repository for the organisation's condition monitoring data. It is driven by *people* using well defined and documented *processes*.
- ***People*** are responsible for the day to day management of the system. They develop and use *processes* to enter and consume data in the *information system*.
- ***Processes*** ensure consistent data entry from all relevant sources of condition monitoring data. The *processes* also ensure that the authorised *people* are properly trained and extract maximum value from the data contained in the *information system*

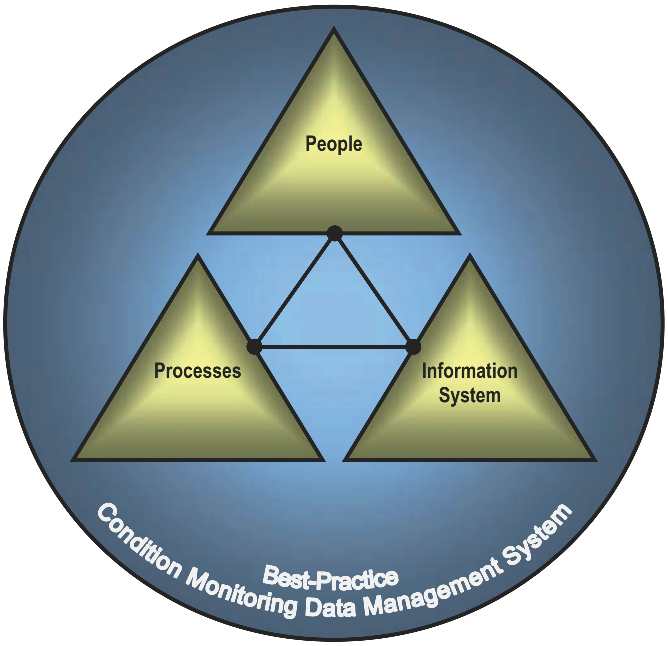


Figure 2: A Best-practice system consists of three equally important and inter-dependent components working in synergy

3.1 Information System

To allow an organisation to realise the true value of condition monitoring data there needs to be a fundamental change to the way that this data is managed. Typically, condition monitoring reports are delivered to the organisation in a non-divisible entity like a Word or PDF document. Instead, the data contained within the reports needs to be "groomed" into a collection of individual records of observed equipment condition, or "Condition Records". Each "Condition Record" is a set of data fields that in most cases gives the maintenance personnel enough information to 1) Understand what test was done and when, on what component and by whom 2) Understand what the interpretation of the observed test results was and 3) Understand what preventative maintenance, if any, is recommended. This is, regardless of the technique used, the results for each individual condition survey should be distilled into a common set of data fields such as

Component Tested, Test Date, Technician, Test Technique, Interpretation of Results and *Maintenance Recommendation(s).* The process of grooming the data ensures data consistency between test techniques and trims the data of all non-essential information. This allows the maintenance personnel to quickly understand the situation without being overwhelmed with unnecessary data. While additional test specific information (a vibration frequency spectrum for example) should be accessible with just a few extra mouse clicks, in most cases it is not required in the day-to-day management of the plant.

The process of grooming the condition monitoring data before it is stored is a key enabling step for achieving best-practice condition monitoring data management. Simultaneously, the underlying storage system for condition monitoring data must be changed from existing ad-hoc systems such as filing cabinets or hard disk storage, to a dedicated (but not necessarily complex) information system. The information system must be capable of storing the groomed condition monitoring data for each asset in the plant. It must also then allow maintenance personnel to query this information, ideally from any location and at any time.

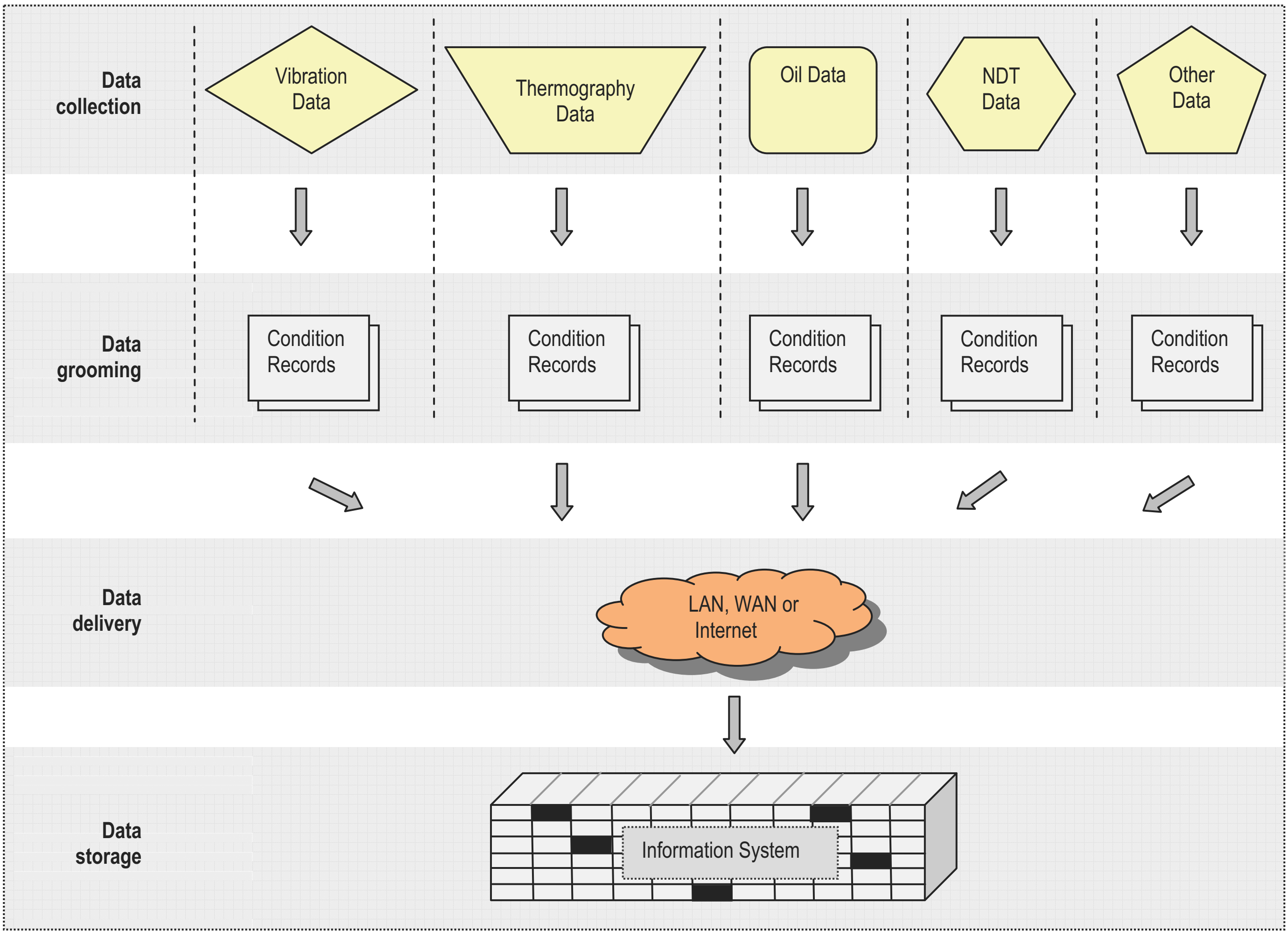


Figure 3: A Best-practice approach to condition monitoring data management requires the data from each technique to be "groomed", that is, broken down into a basic set of common data elements, before being delivered to an information system. This approach enables the data to be easily accessed by everyone in the organisation and also allows for value adding through future data mining.

3.2 People

Making the transition from ad-hoc to best-practice requires the commitment of a number of different groups of people from both inside and outside the organisation. While maintenance personnel will consume the data in the information system on a daily basis, other people such as project managers, expert consultants, external contractors, OEM employees and IT Support staff may all be required to implement the system and to keep it running. All of these people need to understand the potential benefits of a best-practice approach and need to be aligned to the goal of achieving it.

3.3 Processes

A set of well defined process and procedures is required to commission and maintain a best-practice system. Well defined processes will ensure that data integrity and completeness is maintained and that maximum value is leveraged from the collective condition monitoring data. Typical processes might include

- Well documented data formatting and transmission standards to ensure data integrity in the likely case that there is more than one source of condition monitoring data
- Developing training procedures and user manuals to ensure that maintenance personnel extract the most value from the organisation's consolidated condition monitoring data.
- A complete set of IT system management processes to ensure that the underlying information system remains secure, stable and available around the clock.

4 WHAT ARE THE BENEFITS OF BEST-PRACTICE CONDITION MONITORING DATA MANAGEMENT

There are significant direct and indirect benefits for the organisations that achieve best-practice condition monitoring data management.

Silos of data are eliminated: Organisations benefit through instant enterprise-wide access to condition monitoring data. There is only one single, centralized storage location for all data and provided it is an Internet based system, authorized personnel can access the data at any time and from any location. There is minimal effort required to collate condition monitoring data across multiple techniques for a particular asset. Maintenance decisions are more informed and more efficient because all relevant data is considered, rather than just the results of the most recent condition monitoring report for the asset in question. The ability to quickly and easily search for and find all relevant condition monitoring data for an asset could collectively save hundred's of person-hours across the enterprise each year. This is particularly true if the organisation is involved in detailed reliability improvement projects, such as RCM, where a detailed analysis of component history and failure modes may be required. In this case, a detailed condition monitoring history is a good complementary source of failure data to the work order history contained in the CMMS system.

Ability for multi dimensional data analysis: The information system will allow maintenance personnel to effortlessly retrieve a complete condition monitoring history for a particular component. Just as easily, the information system can retrieve alternative views of the data, such as a snapshot of the current health of all assets across the plant. This particular view of the data will result in maintenance activity being better directed towards those assets that need it the most. If a criticality analysis of the plant has been conducted, this information can be combined with the condition data to show a matrix of condition severity versus component criticality. Components that are in poor condition *and* highly critical with respect to risk to safety, environment or production potential are flagged to ensure they receive prompt attention.

Other views of the condition monitoring data may include various exception reports, for example the system could report on any items of equipment that were be inspected for some reason, for example, access to the asset was not granted by the operator. Risk is reduced by identifying those assets where inspections are repeatedly missed, particularly if the asset is highly critical. Another example would be an exception report of outstanding maintenance recommendations. This can help ensure that maintenance personnel acknowledge and act on maintenance recommendations made by the condition monitoring technicians and that nothing falls through the cracks.

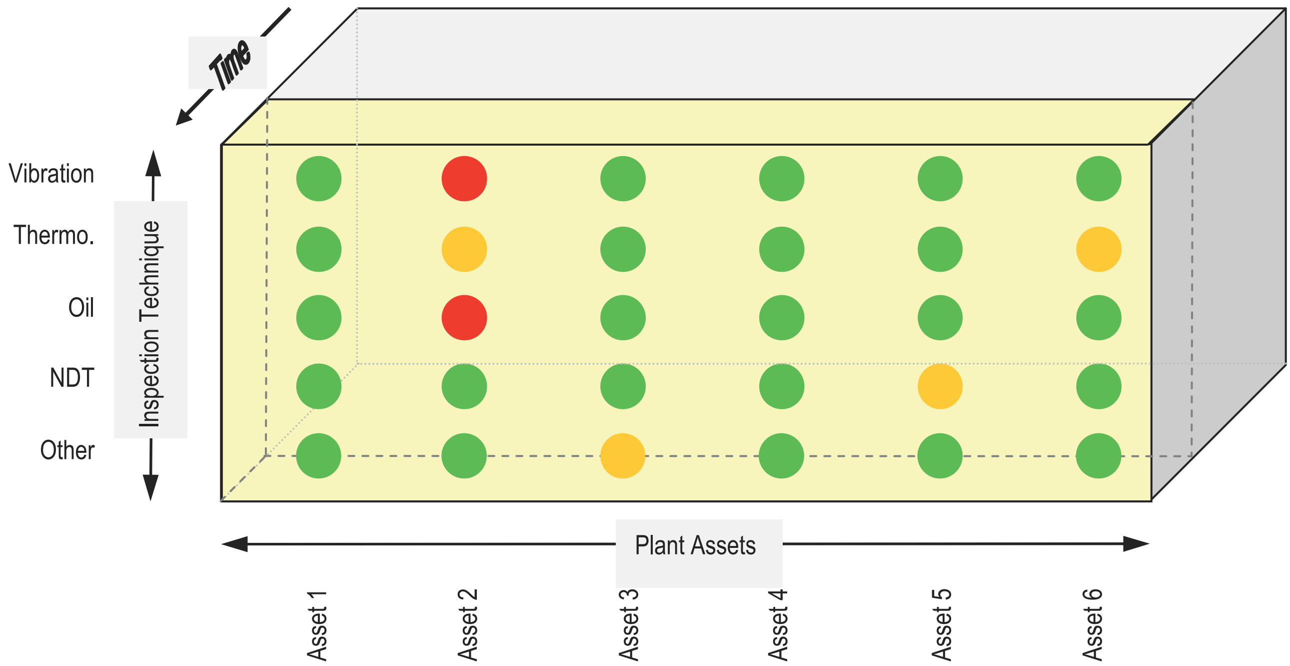


Figure 4: A dedicated information system enables multi-dimensional views of the data. In this example, maintenance personnel can view an up-to-date snapshot of overall plant health. Based on the most recent vibration, thermography and oil surveys, Asset 2 is currently in poor condition and requires maintenance

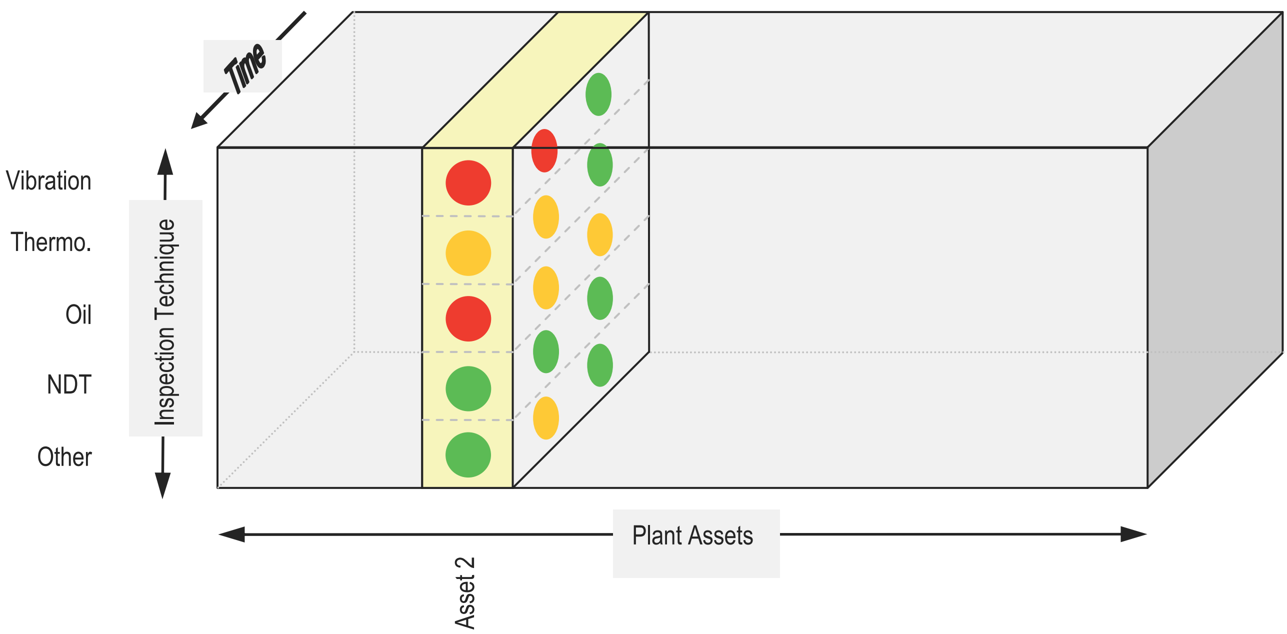


Figure 5: In an alternate view of the data, the user is able to view a complete condition monitoring history across all inspection techniques for Asset 2. This holistic view of the health of Asset 2 enables the maintenance personnel to make the most informed maintenance decisions.

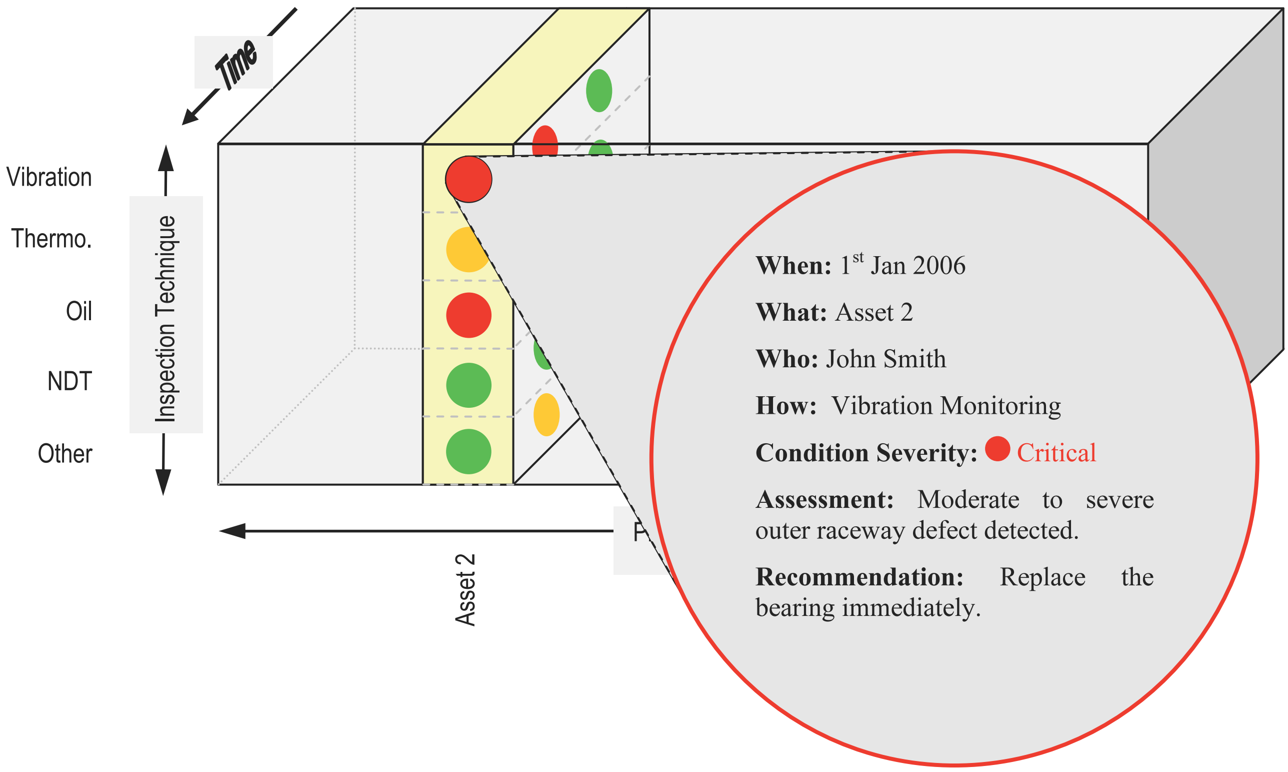


Figure 6: With a few mouse clicks the user can drill down into the individual details of each inspection

Reduced failures through rigorous management of Maintenance Recommendations It becomes easier to actively manage the maintenance recommendations made by the various suppliers of condition monitoring services. The information system can track and flag any assets that have outstanding maintenance recommendations which helps ensure that nothing falls through the cracks. This is also an obvious integration point with the CMMS system. Ideally, maintenance personnel can review the list of outstanding maintenance recommendations on a daily or weekly basis and easily trigger work orders in their CMMS system with a few simple clicks, saving data entry time and increasing data integrity. Effective management of maintenance recommendations also gives important feedback to the condition monitoring technician who needs to know whether previous recommendations were followed or deferred before conducting subsequent surveys on that asset.

Calculating the true value of the condition monitoring program: The organisation can understand and quantify the value of the condition monitoring program by tallying the total savings of avoided downtime as a direct result of the recommendations made by condition monitoring technicians. This can assist the maintenance group justify the annual condition monitoring spend.

Monitoring key performance indicators: Centralising the condition monitoring data enables a number of KPI's to be monitored. An example would be an overall score of plant health. A successful condition monitoring program should result in this score being maintained or improved over time. Different areas of the plant can be benchmarked against each other or a nominated best-practice goal. Another example would be a score of the number of outstanding maintenance recommendations in each area of the plant. This will rank each area on how quickly the maintenance personnel are acting on the recommendation made by the condition monitoring technicians.

Ability for data value adding: The process of grooming the condition monitoring data into a common format enables potentially limitless analysis to be conducted, at the asset, plant or even enterprise level. Over time, the collective value of this data becomes much greater than the sum of its parts, allowing enterprises to convert this data into more valuable knowledge. Consider the following examples

1. *Value add 1: Optimizing the condition monitoring program*: The information system can identify the components that remain in "as new" condition over an extended period of time. The organisation can periodically conduct a detailed analysis of these components and perhaps increase the inspection interval or even remove the component from the condition monitoring program all together.

2. *Value add 2: Enterprise wide query for problem components*: If a problem component is identified within an organisation it becomes easier to find out whether other areas within the organisation or even the enterprise are experiencing similar health problems with that component. This may help identify components where design revision is required or an alternative component should be sourced.

3. *Value add 3: Enterprise wide benchmarking of similar assets:* The long term condition of similar assets across the organisation can be compared. This can highlight maintenance practices that result in better long term asset performance. Conversely, poor or inconsequential maintenance practices can be identified and eliminated.

5 THE ROADMAP – HOW TO MAKE THE TRANSITION FROM AD-HOC TO BEST-PRACTICE CONDITION MONITORING DATA MANAGEMENT

Like any business process improvement project, the path to achieving best-practice management of condition monitoring data can be difficult. Significant challenges can arise in obtaining the support of management and stakeholders as well as coordinating all of the affected parties, particularly if there are multiple sources of condition monitoring data. It is important to begin by conducting a thorough analysis of the existing systems in the organisation versus the desired best-practice. From there, the scope of work becomes clearer and a cost/benefit analysis can be presented to stakeholders. It is likely that a cost/benefit analysis will strongly support the investment in a best-practice approach to condition monitoring data management.

A suggested approach to the process improvement project is outlined below

1. *Choose a best-practice champion:* Nominate someone within the organisation, or even a third party, to take responsibility for spearheading the push towards best-practice data management. The champion should be able to demonstrate the following experience and capabilities.
    a. Familiarity with data management and information systems.
    b. A basic understanding of the technology and processes used for each of the condition monitoring techniques
    c. A broad understanding of the organisation and all sources of condition monitoring data (both internal and external)
    d. Excellent communication and project management skills
    e. The champion should have the authority to commandeer the required resources to keep the project moving. This may include OEM's, in house and contract Condition Monitoring suppliers, maintenance personnel, management and other support personnel (IT Support Staff for example)

2. *Map the current business processes in the context of condition monitoring data:* Conduct an analysis of all sources of condition monitoring data across the organisation. This could include in-house technicians, OEM's, condition monitoring contractors and permanent on-line systems. For each source of data document the data format and how it is delivered to the maintenance personnel. Also document how the data is then permanently stored and where; filing cabinets, network hard drives, information systems, etc.

3. *Map the desired business process:* Identify and map the ideal business processes that would allow the organisation to achieve the benefits of best-practice condition monitoring data management. In particular, consider the ideal information flow between the condition monitoring technicians and the consumers of the data.

4. *Determine the functional requirements of the information system:* List the functions that an ideal information system will perform. Consider the various users of the system and the way in which they need to consume the data. For example, a maintenance manager might just need to see a rolled up summary of current plant health each day while a planner may need to see a holistic view of the condition monitoring data for an asset. Can the existing CMMS system be customised to meet these requirements? Is there an existing off-the-shelf solution that can be adapted or is there are need to develop a dedicated system?

5. *Decide on the people required to implement the desired business process:* Achieving the change to best-practice analysis above will require the ongoing efforts of a number of people, both inside and outside the organisation. Identify which people will be affected and how. If necessary, precisely document the changes that will be required from these people and the ongoing responsibilities and expectations moving forward.

6. *Decide on the processes required to reach the desired business process:* A new set of processes will be required to ensure that entire system operates effectively on a day-to-day basis. The data format and data entry and transmission expectations need to be documented. New training processes will need to be developed and delivered to personnel who supply condition monitoring data to the system. Training will also be required for the maintenance personnel who consume the data on a day-to-day basis.

7. *Conduct a gap analysis:* By comparing the current verses the desired business processes, determine the scope of work required to bridge the gap.

8. *Formulate a plan to implement the required changes to the processes, people and information systems*. At this stage the scope of work should be clearer and a work breakdown can be formulated. From this work breakdown the cost and duration of the entire process improvement project can be estimated.

9. *Conduct a cost/benefit analysis*: Compare the estimated cost of the project to the potential long term benefits. To estimate the financial benefits of a best-practice approach, consider the ongoing savings through
    a. Eliminated failures due to better integration and communication of condition monitoring data. For this step it may help to analyse a number of previous failures to determine whether a best-practice approach to data management may have eliminated the failure or reduced it's financial impact.
    b. Time saved through instant access to integrated condition monitoring data across the enterprise
    c. Potential benefits of the ability to convert data to knowledge through data mining across the enterprise.

    In many organisations, the cost benefit analysis will show a payback on investment well within the first few years. This detailed cost/benefit analysis can be presented to the organisation stakeholders and a go/no-go decision can be made.

10. *Formalise a project plan and communicate the plan to all affected parties:* Provided that stakeholders are now committed to the project, the work breakdown can be more formally set into a project plan. The plan should set some specific implementation goals and deadlines. It will also show the requirements for labour and financial resources to allow the project to be accurately budgeted. Use the plan and other supporting material to communicate to all effected parties
    a. The objectives of the process improvement project
    b. The anticipated benefits of moving to a best-practice approach
    c. How they will be effected during the implementation phase of the project and what the ongoing expectations are
    d. Where the project deadlines and milestones are.
    e.

5.1 The Pitfalls and Challenges

There are a number of pitfalls and challenges that may arise when improving the condition monitoring data management processes. Some of these challenges are identified and risk mitigation strategies are suggested.

Lack of stakeholder support: It is vital to the success of the project that adequate resources are allocated and there is full support from stakeholders. To ensure this is the case it is critical that a detailed cost / benefit analysis is conducted and presented to the stakeholders so that the true value of the process improvement project is understood from the start. This is not a process improvement project that can be delegated to an employee to be undertaken in their "spare time". It requires dedicated resources and the cooperation of a number of parties both inside and outside the organisation.

Relying on the CMMS to manage condition monitoring data: Typically, existing CMMS systems are not ideally suited to managing large volumes of condition monitoring information in the manner described above. To do so the CMMS will likely require some customization, which may be both time consuming and expensive. A better solution may be to adopt or custom develop a dedicated information system specifically for this task. The chosen information system may be integrated with the CMMS system where logical, for example, progressing a maintenance recommendation into a CMMS work order.

Difficulties in integrating multiple sources of condition monitoring information: In some organisations there will be many sources of condition monitoring data. This can make it difficult to "groom" the condition monitoring data into a common format. One way to mitigate this risk would be to narrow the scope of the initial project to a smaller part of the overall condition monitoring program. For example, an organisation may choose to integrate only oil and vibration monitoring data to begin with. This will reduce the initial cost and risk of the project and allow teething issues to be rectified before the scope is widened to include the remaining condition monitoring techniques. Another way to mitigate this risk would be to consolidate the suppliers of condition monitoring data across the organisation, hence reducing the sources of data. Once the system is in place, it is also vital to ensure that all suppliers of condition monitoring data (both internal and external) strictly adhere to the established data entry processes to ensure data integrity and the completeness of the data. This again highlights the importance of well defined and well communicated processes.

6 CONCLUSION

Best-practice management of condition monitoring data isn't itself a solution to overall plant reliability. However, it is an essential part of the wider maintenance program. It is difficult for organisations to achieve maintenance program maturity when there is an ad-hoc approach to storage and management of condition monitoring data. With a robust data management system in place organisations save significant time and money in the short term through quick and easy enterprise wide access to the data and through more informed and timely maintenance decisions. In the long term these organisations are able to convert this data into knowledge to optimize the predictive maintenance program and improve overall plant reliability.

OPTIMIZATION OF INTENTIONAL MISTUNING FOR BLADED DISK : INTENTIONAL MISTUNING INTENSITY EFFECT

ByeongKeun Choi[a], Kim Hack Eun[c], Kim Hyo Jung[b], Jeong Haneol[b], Gu DongSik[b], Kim Myung Kwan[c] Author

[a]*School of Mech. and Aerospace Eng., Gyeongsang National Univ., Tongyoung, Kyongnam 650-160, Korea*

[b]*Precision Mech. Eng., Gyeongsang National Univ., Tongyoung, Kyongnam 650-160, Korea*

cKorea Gas Technology Corporation, #1179, Anjeong-Ri, Gwangdo-Myeon, Tongyoung, Gyeongnam, 650-824, Korea

Abstract: In turbomachinery rotor, there are small differences in the structural and/or geometrical properties of individual blades, which are referred to as blade mistuning. Mistuning effects of the forced response of bladed disks can be extremely large as often reported in many studies. In this paper, the pattern optimization of intentional mistuning for bladed disks considering with intentional mistuning intensity effect is the focus of the present investigation. More specifically, the class of intentionally mistuned disks considered here is limited, for cost reasons, to arrangements of two types of blades (A and B, say) and Genetic Algorithm is used to optimize the arrangement of these blades around the disk to reduce the forced response of blade with intentional mistuning intensity levels.

Key wards : Intentional Mistuning, Mistuning Intensity, Bladed Disk, Pattern Optimization

1 INTRODUCTION

In a dynamic analysis of a turbomachinery rotor, one traditionally has assumed that the blades are identical. But, in practice there are small differences in the structural and/or geometrical properties of individual blades, which are referred to as blade mistuning. Much of the vast literature on this topic [1-5] has assumed these differences to be small and to arise either during the manufacturing process and/or as a consequence of in service wear. The motivation for considering such small variations is that their effects of the forced response of bladed disks can be extremely large as often reported in the above studies. Interestingly, the large sensitivity of the tuned system to these small variations has been linked [3] to its high level of symmetry.

In this light, it would appear beneficial to design bladed disk not to be tuned, namely to exhibit intentional mistuning, to reduce the sensitivity of the forced response to unintentional mistuning. Certainly, the consideration of intentional mistuning is not new [1,2] however, in the context of forced response, some papers [6,7] have only recently investigated the use of harmonic patterns of mistuning. Recently, the authors have investigated and identified the effect of intentional mistuning which can significantly reduce the magnification of the forced response due to unintentional random mistuning. [8,9]

In this paper, the pattern optimization of intentional mistuning for bladed disks considering with intentional mistuning intensity effect is the focus of the present investigation using the two sets of blades A and B. Genetic Algorithm is used to obtain the pattern(s) that yields small/the smallest value of the largest amplitude of response to a given excitation in the absence of unintentional mistuning using simple model (one-degree-of-freedom per blade) of bladed disks.

2 OPTIMIZATION APPROACH

In view of the complexity and cost of intentional mistuning, one should not look simply at set patterns, for example the harmonic patterns [7] but rather one should optimize the pattern to reduce as much as possible the amplification of the forced response to a given excitation or set thereof. Accordingly, it was suggested by authors that the use of intentional mistuning is probably not a standard design tool but would be very valuable if: (a) it yields a large decrease in sensitivity to unintentional mistuning, and (b) it involves a minimum number of types of blades, ideally 2. [8,9]

Therefore, in this paper, the disk will first be assumed to support only two different types of blades (blades A and B, say) and their arrangement that yields the smallest amplification of the forced response will be sought with different intentional mistuning intensity. Many sets of blades A and B were selected to change the intentional mistuning intensity having natural frequencies lower and higher, respectively, than the tuned ones (type C) in this paper.

Mistuning also produces a very nonlinear effect on the forced response. That is, by switching the order of the blades around the disk, dramatic differences can be obtained in the variability of the blade-to-blade amplitudes of vibration as exemplified in particular by the harmonic mistuning analysis of Mignolet et al. [4]. It might thus be suspected that there exists a series of local optima in the complex, high dimensional space over which the optimization must take place. In this light, the present optimization effort has relied on the use of Genetic Algorithm (GA) [10].

3 SIMPLE GENETIC ALGORITHM

GA are particularly well suited for the present effort because the design variables only admit discrete values (i.e. a specific blade is only of type A or B), see References [10] for further details. The simple genetic algorithm (SGA) used here relies on a population of n_{pop} bladed disks each of which is a random arrangement of N genes (the type A or B of the different blades). Accordingly, each bladed disk can be characterized by a sequence of N A and B letters, for example AABBBBAA…, which evolves from one generation to the next according to the rules of selection, crossover, and mutation until all the chromosomes yield essentially similar values of the fitness or objective function(the maximum amplitude of blade response).

The fitness proportionate selection, the single point crossover technique and an exponentially decreasing mutation function was used in the present investigation. Also, the one elite reservation strategy was used in this paper according to which the best disk is retained unchanged from one generation to the next.

4 SIMPLE BLADED DISK MODEL

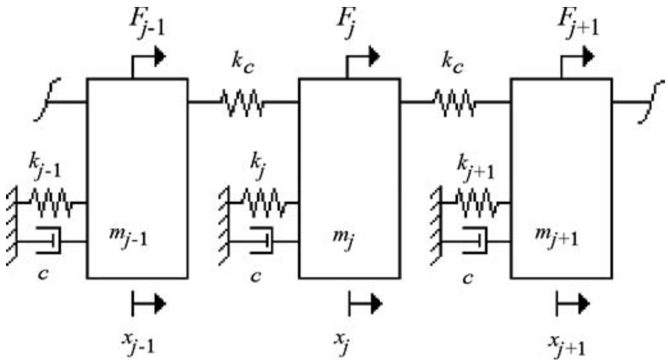


Figure. 1 Single-degree-of-freedom per blade disk model

$$F_j = F_o \cos\left(\omega t + \frac{2\pi r (j-1)}{N}\right)$$

$m_t = 0.0114$ kg, $k_t = 430{,}300$ N/m, $c = 1.3814$ N·s/m, $N = 4$ blades, $F_0 = 1$N, $r = 3$,

k_j : normal distribution of mean k_t and standard deviation σ

To identify the effect of intentional mistuning intensity to optimum pattern of intentional mistuning, the SDOF per blade model shown in Fig. 1 was first considered with the coupling stiffness (k_c = 45430 N/m) and damping coefficient ($c = 1.3814$ N s/m) with only 4 blades was considered in this paper. Specifically, each of the N blades is represented as a single mass (m) which is connected to the ground (i.e. the disk) and the aerodynamic and structural coupling between blades are modeled by springs (k_c) and dashpots (c). In the sequel, it is assumed that the coefficient vanishes in this paper. The values of mass (m = 0.0114 kg), and stiffness (k_t = 430,300 N/m) have already been used in a previous investigations to model a high-pressure turbine stage is used in this paper.

The computations proceeded as follows. The bladed disk model and engine order of the excitation were first selected. The SGA described above was then used to obtain the intentionally mistuned disk formed of blades A and B such that the maximum of its response over the entire frequency range was the smallest possible.

First, the intentional mistuning intensity to optimum pattern of intentional mistuning is identified. Many sets of blades A and B were selected to change natural frequencies lower and higher, respectively, than the tuned ones (type C). In analysis, the stiffness of A and B type blade is selected to have lower and higher than the tuned ones. For example, in table 1, A(0.9) and B(1.1) means that A and B have natural frequencies 5% lower and 5% higher than C. The optimum pattern of intentional mistuning is searched with 45,430 N/m (an average to strong blade-to-blade coupling level) by Genetic Algorithm.

Table 1 show the comparison of optimization result by Genetic Algorithm for each intentional mistuning intensity level with the tune system (all C:1.0682e-4 m and all B) in 3th engine order case.

Table 1 The comparison of optimization result by Genetic Algorithm for each intentional mistuning blade set with the tune and other mistuning pattern

Mistuning blade type		Optimum distribution
A	B	
0.7	1.3	BBBB = 9.5817e-5 m (ABBA = 1.0026e-4 m)
0.8	1.2	ABBA = 9.7014e-5 m, (BBBB = 9.9172e-5 m), 9.18%
0.9	1.1	ABBA = 9.9242e-5 m, (BBBB = 1.028e-4 m), 7.09%
0.91	1.09	ABBA = 1.0186e-4 m, (BBBB = 1.0326e-4 m)
0.9125	1.0875	ABBA = 1.0261e-4 m, (BBBB = 1.0305e-4 m)
0.92	1.08	BBBB = 1.0344e-4 m, (ABBA = 1.0536e-4 m)
0.94	1.06	BBBB = 1.0425e-4 m, (ABBA = 1.163e-4 m)
0.96	1.04	BBBB = 1.0529e-4 m, (ABBA = 1.2457e-4 m)
0.98	1.02	BBBB = 1.0619e-4 m, (ABBA = 1.1438e-4 m)

In Table 1, the optimum pattern that yields small/the smallest value of the largest amplitude of response to a given excitation by GA was changed according to the intentional mistuning intensity (mistuning blade type) level. The genetic optimization algorithm yielded the configuration ABBA and BBBB the highest responding blade in each mistuning blade type. Fig. 2 ~ 4 shows the forced response of optimum, tuned and other A/B patterns in three mistuning type. It is identified that ABBA are better then tuned and other patterns in cases of A(0.9), B(1.1) and A(0.9125), B(1.0875). But in case of A(0.94) and B(1.06), BBBB is better than tuned and ABBA.

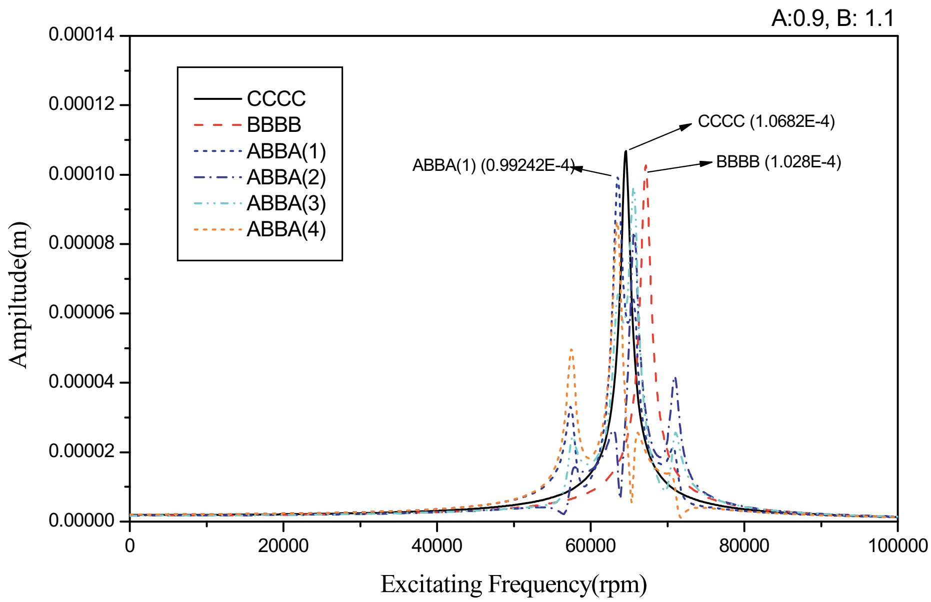


Figure. 2 Comparison with forced response of optimum, tuned and other A/B pattern(A:0.9, B:1.1)

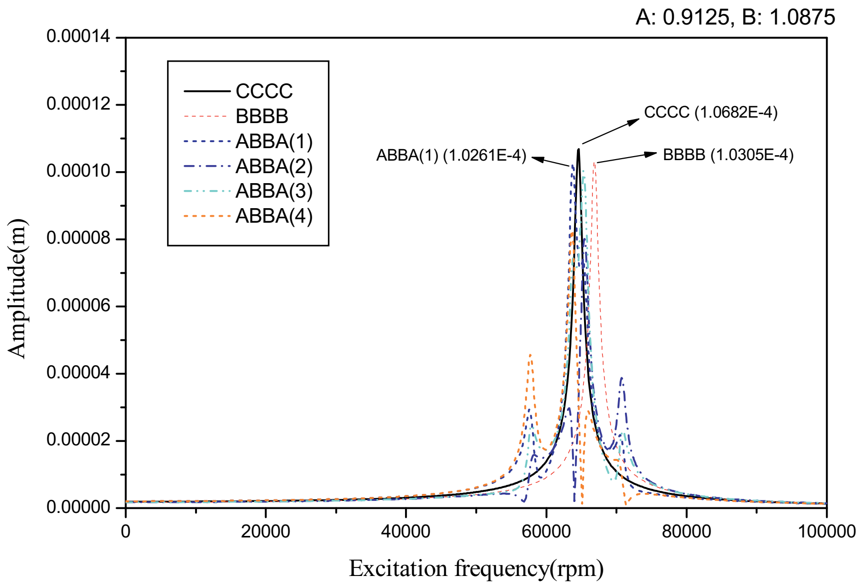


Figure. 3 Comparison with forced response of optimum, tuned and other A/B pattern(A:0.9125, B:1.0875)

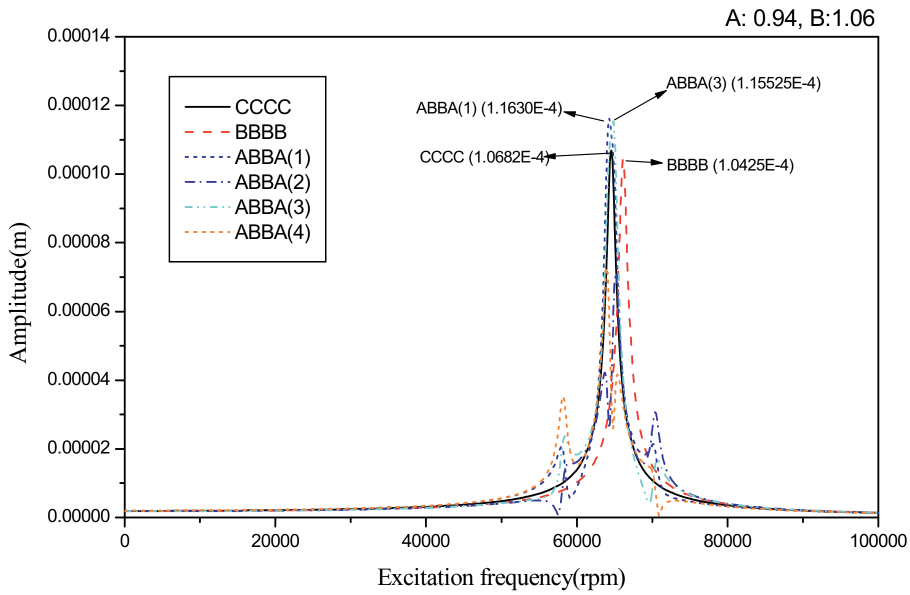


Figure. 4 Comparison with forced response of optimum, tuned and other A/B pattern(A:0.94, B:1.06)

According to Table 1, in A(0.9), B(1.1) and A(0.8), B(1.2), the max amplitude of blade is reduced to 9.18% and 7.09% than tuned one respectively. So, in this paper, more detail analysis around the A(0.9) and B(1.1) is needed to identify the effect of intentional mistuning intensity to optimum pattern of intentional mistuning. Table 2 show the comparison of optimization result for intentional mistuning intensity levels around A(0.9) and B(1.1) with the tune system (all C and all B) in 3th engine order case.

In Table 2, as the gab between A and B is larger, all A become worse; max amplitude is increased while all B and ABBA become good; max amplitude is decreased. Also, if the gab between A and B is equal, the larger A or the larger B is better than others. That means that the max amplitude is more reduced than tuned one.

Table 2 The comparison of optimization result for intentional mistuning intensity with the tune and ABBA

A	B	B - A	All A	All B	ABBA	Compare to all B	Compare to all C
0.9	1.1	0.2	1.1155e-4 m	1.028e-4 m	9.9242e-5 m	3.46%	7.09%
0.9 + 0.01(0.91)	1.1 + 0.01(1.11)	0.2	1.1124e-4 m	1.0238e-4 m	9.8977e-5 m	3.32%	7.34%
0.9 – 0.01(0.89)	1.1 – 0.01(1.09)	0.2	1.1201e-4 m	1.0326e-4 m	9.9536e-5 m	3.6%	6.82%
0.9 + 0.01(0.91)	1.1 – 0.01(1.09)	0.18	1.1124e-4 m	1.0326e-4 m	1.0186e-4 m	1.36%	4.64%
0.9 – 0.01(0.89)	1.1 + 0.01(1.11)	0.22	1.1201e-4 m	1.0238e-4 m	9.7272e-5 m	4.99%	8.94%
0.9 + 0.02(0.92)	1.1 + 0.02(1.12)	0.2	1.1078e-4 m	1.0209e-4 m	9.8468e-5 m	3.55%	7.82%
0.9 + 0.02(0.92)	1.1 – 0.02(1.08)	0.16	1.1078e-4 m	1.0344e-4 m	1.0536e-4 m	-1.86%	1.37%
0.9 – 0.02(0.88)	1.1 + 0.02(1.12)	0.24	1.1273e-4 m	1.0209e-4 m	9.5777e-5 m	6.18%	10.34%
0.9 – 0.03(0.87)	1.1 + 0.03(1.13)	0.26	1.1332e-4 m	1.0153e-4 m	9.4605e-5 m	6.82%	11.44%
0.9 + 0.008(0.908)	1.1 + 0.008(1.108)	0.2	1.1112e-4 m	1.0252e-4 m	9.8741e-5 m	3.69%	7.56%
0.9 – 0.012(0.888)	1.1 – 0.012(1.088)	0.2	1.1239e-4 m	1.032e-4 m	9.9855e-5 m	3.24%	6.52%
0.9 + 0.01(0.91)	1.1 – 0.008(1.092)	0.182	1.1124e-4 m	1.0307e-4 m	1.0147e-4 m	1.55%	5.0%
0.9 – 0.008(0.892)	1.1 + 0.01(1.11)	0.218	1.1215e-4 m	1.0238e-4 m	9.7473e-5 m	4.79%	8.75%
0.9 – 0.01(0.89)	1.1 + 0.008(1.108)	0.218	1.1201e-4 m	1.0252e-4 m	9.7541e-5 m	4.86%	8.69%

5 SUMMARY

The investigation of this paper focused on the pattern optimization of intentional mistuning for bladed disks considering with intentional mistuning intensity effect is the focus of the present investigation using the two sets of blades A and B. Genetic Algorithm is used to obtain the pattern(s) that yields small/the smallest value of the largest amplitude of response to a given excitation in the absence of unintentional mistuning using simple model (one-degree-of-freedom per blade) of bladed disks.

Through the optimization, it is found that the optimum pattern that yields small/the smallest value of the largest amplitude of response to a given excitation by GA was changed according to the intentional mistuning intensity (mistuning blade type) level. The genetic optimization algorithm yielded the configuration ABBA and BBBB the highest responding blade in each mistuning blade type in Table 1. Also it is identified that the gab between A and B is larger, all A become worse(namely, max amplitude is increased) while all B and ABBA become better(namely, max amplitude is decreased). If the gab between A and B is equal, the larger A or the larger B is better than others, namely, the max amplitude of blade is more reduced than tuned one in Table 2. Therefore, the effect of intentional mistuning intensity should be considered to optimize the intentional mistuning pattern which can reduce the forced response in blade.

Acknowledgements

This study is supported by Second-Phase of BK21 Project and New University for Regional Innovation (NURI) Project.

6 REFERENCES

[1] Kaza, K.R.V., and Kielb, R.E. (1984) Flutter of Turbofan Rotors with Mistuned Blades. *AIAA Journal*, 22(11), 1618-1625.

[2] Kaza, K.R.V., and Kielb, R.E. (1985) Vibration and Flutter of Mistuned Bladed-Disk Assemblies. *Journal of Propulsion*, 1(5), 336-344.

[3] Wei, S. T., and Pierre, C. (1988) Localization Phenomena in Mistuned Assemblies with Cyclic Symmetry - Part I: Free Vibrations. *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, 110(4), 429-438.

[4] Mignolet, M.P., Hu, W., and Jadic I. (2000) On the Forced Response of Harmonically and Partially Mistuned Bladed Disks. Part I: Harmonic Mistuning. *International Journal of Rotating Machinery*, 6(1), 29-41.

[5] Mignolet, M.P., Hu, W., and Jadic I. (2000) On the Forced Response of Harmonically and Partially Mistuned Bladed Disks. Part II: Partial Mistuning and Applications. *International Journal of Rotating Machinery*, 6(1), 43-56.

[6] Castanier, M.P., and Pierre, C. (1997) Predicting Localization via Lyapunov Exponent Statistics. *Journal of Sound and Vibration*, 203(1), 151-157.

[7] Castanier, M.P., and Pierre, C. (1998) Investigation of the Combined Effects of Intentional and Random Mistuning on the Forced Response of Bladed Disks, *Paper AIAA-98-3720*.

[8] B. K. Choi (2003) Pattern Optimization of Intentional Blade Mistuning for the Reduction of the Forced Response Using Genetic Algorithm. *KSME International Journal*, 17(7), 966-977

[9] B. K. Choi, J. Lentz, A. J. Rivas-Guerra, and M. P. Mignolet (2003) Optimization of Intentional Mistuning Patterns for the Reduction of the Forced Response Effects of Unintentional Mistuning: Formulation and Assessment. *ASME Journal of Engineering for Gas Turbines and Power*, 125(1), 131-140.

[10] Gen, M., and Cheng, R. (1997) Genetic Algorithms & Engineering Design. New York, Wiley-Interscience.

IDENTIFYING AND MANAGING THE COMMUNITY INFRASTRUCTURE ASSET RENEWAL GAP

J Hennessy [a] and DG Platt [b]

a Municipal Association of Victoria, Melbourne, Victoria 3000, Australia.

b Anstra, Melbourne, Victoria, Australia.

Abstract: Sustaining our community infrastructure and assets is potentially the greatest challenge faced by Local Governments across Australia. At risk is the continued sustainability of assets which support the delivery of many of our community services. Most states have identified a substantial emerging gap in Local Government's capacity to renew it assets. We suggest the problem is broader than the conventional asset management activities and fundamentally requires that we have informed debate at both the national and local levels. In order to have such an informed debate we need access to reliable, consistent and relevant information. However, experience has shown that trying to obtain this information is difficult. In most cases it does not exist and where it does exist it is frequently inadequate. This paper outlines a project that is currently in progress that aims to develop a national framework for the exchange of community asset information. As a result we hope to be able to inform the debate that must occur if we are to solve the issue of the renewal gap for community assets.

Key Words: local government, renewal gap, community, assets, reference model, information framework

1 INTRODUCTION

Sustaining our community infrastructure and assets is probably the greatest challenge faced by Local Governments across Australia. The issue is not asset management as an end game, rather the continued sustainability of the communities in which we live, work and recreate. To put the problem in perspective Local Government is responsible for 81% of the nation's roads, has over $154 billion of assets, of which roads are 64% [1] and in 2003/4 spent over $3.8 billion on local roads [2]. Field observations suggest that a large proportion of this infrastructure is rapidly approaching the end of its useful-life, with a large gap emerging between our ability to sustain the renewal of these assets into the future. Whilst the astute have been able to foresee the problem it is only recently that governments have started to systematically assess the situation and quantify the problem. Over the last decade the states have each initiated projects that will help them better understand the true extent of the problem [3].

In Victoria the process of developing a shared understanding across all councils took its first major step through a review of asset conditions and asset management practices in 1997-1998, "Facing the Renewal Challenge"[4]. The outcome of the review was that for the first time it was possible to forecast community infrastructure and asset renewal requirements across the State using quantifiable and consistent measures that looked beyond the immediate budget cycles.

The Municipal Association of Victoria (MAV) has responded to this challenge through its development of the Step Program. The Step program has four key pillars, policy, strategy, plans and operations and through these four streams seeks to continue developing the collective understanding of the challenges faced by councils to develop realistic and achievable outcomes for their communities.

Through the Step Program the MAV has been able to develop a better understanding of the Renewal Gap for the various asset classes. A summary of these results showing the capital funding gap over various time horizons is shown in figure 1.

Evidence from programs, such as Step, demonstrate that the renewal gap is real. The issue is how to face the challenge. Continued professional development of asset management is core but in-itself is probably insufficient. We need to understand the purpose served by community assets; transforming the traditionally narrow focus on activities associated with asset management to the outcomes that community infrastructure and assets enable.

Ultimately, the purpose of renewing assets is not to preserve them but to enable communities to achieve the outcomes they belief are important to them and potentially their children. Within this context the debate about the renewal gap should not be restricted to a small subset of asset managers. Managing the renewal gap is an issue that goes to the heart of our aspirations for a safe thriving and prosperous community in which we learn, work, rest and play. Decisions about how we manage the gap will inevitably impact our vision for the community in which we live.

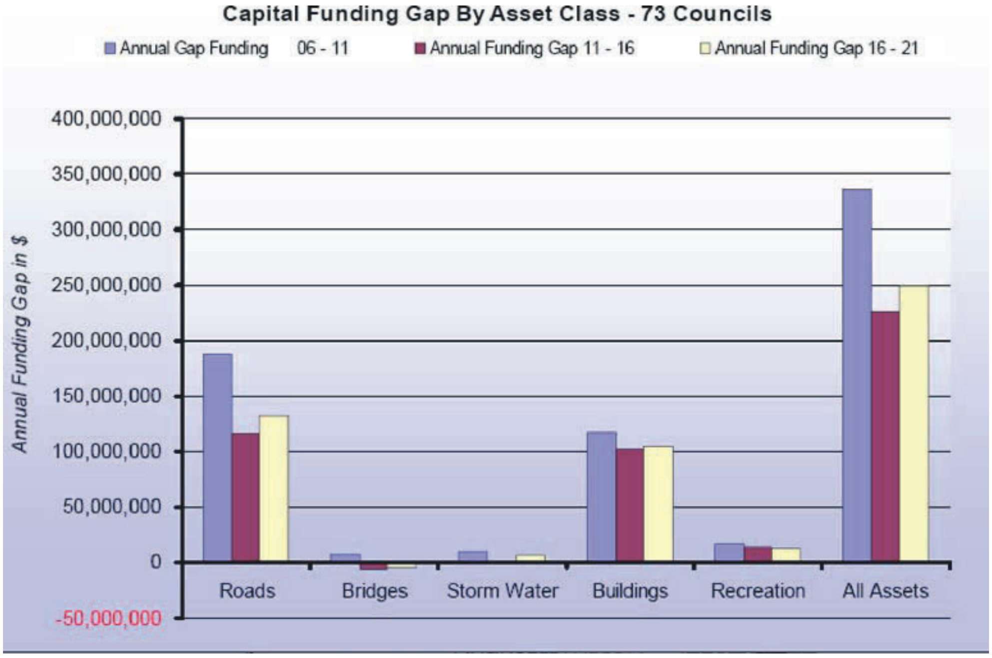


Figure 1. MAV Step Program Capital Funding Gap [5]

In order for the debate to be meaningful it must be supported by sound information. Communities must have access to the right levels of information so as to be empowered to make knowledgeable, even wise decisions about their future. To limit the depth of information or provide superficial summaries is to remove the right from our communities to participate in planning their future. The gap is real and without innovative thinking and potentially some significant changes to the way we manage them the gap will continue.

Understanding why we have a funding gap is important, not least so that we can learn from history. However, it is not the purpose of this paper to explore these issues, this has been done by others [4]. We suggest that the most appropriate way to manage this funding gap is through engagement and empowerment to make sound sustainable decision at both the local and National level.

2 NATIONAL DEBATE

The evidence collected through state programs such as Step [4] suggests that the problem is universal. However, in order to develop a well-founded debate at the commonwealth level we need collectively to aggregate the evidence to present a national perspective. Unfortunately, achieving this has been very difficult as there is no universally agreed core data set or commonality in the definitions or standards used to report information.

Engaging the Federal Government in a debate about innovative ways to bridge the gap in local resources required to sustain the well-being of our communities will require demonstrable evidence of a consistent national problem. In today's political environment there are many other issues competing for the attention and investment of federal resources, many of which have a higher profile in the public arena, such as emergency management, homeland security, refugees and energy and greenhouse emissions etc. The management of community assets and infrastructure has been delegated to Local Government. Local Governments sustain these assets through local rates and grants received from the Federal and State Governments. The impetus to change the status quo lies with Local Government; it must substantiate its case that the level of well-being we believe is appropriate for Australian citizens is not sustainable through the existing arrangements. Moreover, the onus is on Local Government to propose innovative alternatives that demonstrate sustainable and achievable outcomes, anything less is to abdicate the responsibility that has been delegated to it. However, the problem is a collective problem, it is not one that either Local Government or Federal Governments can solve in isolation. The issues are made complex by their interdependencies; robbing Peter to pay Paul is unlikely to provide a sustainable solution, we need to take a holistic approach, which is grounded in truth rather than opinion and myth.

To do this we need reliable, consistent and relevant data; data that can be aggregated to the state and national level yet substantiated down to a local level. To date the ability to generate this quality of information has remained illusive.

3 LOCAL DEBATE

The national debate is primarily focused on the redistribution of resources. The debate within local communities also focuses on the redistribution of resources but in a quite different way. The local community must grapple with the finality of resources. They must understand the potential impact caused by a change in the level of asset preservation on the outcomes and well-being of the whole community.

A realization that under the existing arrangements the sustainability of the outcomes supported by community assets is no-longer viable should trigger a double loop learning debate. The first loop looks at the practices of asset management; can we improve, can we close the gap – there is considerable onus on the professional practitioner to demonstrate improvement. The second loop looks at why we need the asset – the debate must explore the aspirations of the community, understanding the well-beings of our community:

- physical well-being – safe, efficient workable healthy
- economic well-being – thriving and prosperous local economy
- social cultural and spiritual well-being – balance, diversity, amenity, beliefs (not always easily measured as it goes beyond the tangible but often conspicuous by its absence)
- environmental well-being - sustainability, footprints, aesthetics, clean air, clean seas, healthy beaches etc.

It is the second loop that sets the terms of reference in which the professional asset manager operates. The problem is not simple. You can seldom plan to change the level at which an asset is preserved without consideration of several interdependent systems. Consider a drain, and the impact on flooding, which impacts the economic and emotional hardship endured, the value of property the attractiveness of the area or the health of the area. There are many interconnected links between the asset, the services it supports and the outcomes delivered to the community. If the level of service provided by the asset has to change then invariably the outcomes in the community must also change. This is a debate that ought to occur within the community, it may be facilitated by the professional asset manager but we argue it is ultimately a decision that must be made by the local community.

4 INTERDEPENDENT SYSTEMS

The need for asset information is not confined to the asset manager. Making our communities better, safer and more sustainable places requires innovative planning and informed decision making with many different and diverse organisations having a role to play. Many of these organisations are dependent upon Local Government’s services and asset knowledge. Many need access to timely information regarding the service/asset.

Not only are many of these systems dependent upon the services provided by community assets, but they are each interdependent upon the other, a change in the carrying capacity of a bridge structure can impact an emergency management plan, or visa versa a change in the type of proposed response vehicle could have an impact on the level at which the bridge asset needs to be preserved.

The commissions that define the purpose of these organisations are varied, ranging from the planning of integrated transport networks, to the management of counter terrorism for events such as APEC, or mitigation planning for the next bush fire or flood. Arising from their unique commissions each organisation has its own needs and more importantly its own view of the world, which usually results in their having a different view as to how the information should be structured. These individual views of the world are what make the exchange of information difficult. Currently, there is no universal language for the exchange of asset data or the development of consistent asset processes. However, it is our conjecture that their motivations are more tightly coupled, in that most organisations are motivated by the need to ensure the sustainable well-being of the communities they serve. It is this uniting factor that gives us hope that we can develop a shared information framework to support local and national dialogues.

5 FACILITATING GOOD DECISION MAKING

Currently, there are no agreed reference frameworks that consistently describe community infrastructure and assets, such as the road or drainage networks. Organisations that provide services across boundaries are faced with the challenge of aggregating a diverse range of municipal data. Likewise, when councils provide information to 3rd party agencies they find themselves undertaking a massive amount of duplicated effort as each agency describes the problem from their perspective.

As a result the problem is often deemed, too hard and decisions are based on perception not reality. Moreover the lack of information generally means that the only person capable of making an informed decision is the asset manager who is able to draw upon experience and tacit knowledge. However, there is no guarantee that the motivation that steers the heuristic decision making process is aligned with community needs. Furthermore, if we are to solve the issues raised by the renewal gap we will need to be innovative, more of the same will simply exasperate the problem.

In nature the most innovative environments are often the interfaces between systems such as the sea shore, the river bank, the forest edge. Applying the analogy to our situation we suggest that the best way to foster innovation in managing assets is to encourage the interface between asset managers and the multiple communities that have an interest in the outcome; the well-being of our community. We anticipate that the best opportunity for innovative solutions to our problems will arise from the diversity of interaction between communities.

In order to facilitate this interactivity we need to empower our participants with sound information. We need to ensure that the information is consistent at all levels of debate, from local through to national.

6 WHERE ARE WE?

We have got to the point where:

a renewal gap can be substantiated at the state level,

we recognise the need for Local Government to develop a case at the national level,

we understand the need to go beyond traditional asset management activities to manage outcomes,

there is a growing appreciation of the interdependence on shared information between agencies,

solutions to our problems will require innovation and collaboration across communities.

In order to progress there is substantial evidence to suggest that the major impediment is the lack of quality data [7], [8]. This impediment is expressed strongly in a recent review of the commonwealth road funding which was tasked with recommending a revised formula for the interstate distribution of the local road grants based on an assessment of the relative needs of the States for expenditure on maintenance and preservation of local roads. The review found that

"The data available on local road expenditure, local road use and the inventory of local roads and bridges to undertake the assessment of maintenance expenditure on local roads are far from ideal. This indicates that judgments about the most appropriate assessment methods need to balance conceptual considerations against considerations of the consistency and reliability of the available data.

The inadequacy of data that is critical to this Review raises concerns about how well we will be able to satisfactorily fulfil the requirements of the terms of reference." [7]

7 INTEGRATED INFORMATION FRAMEWORK

We need an integrated information framework, one that enables us to address the information needs at community, state and national levels. We don't want three more layers of information management. We must stop solving our problems in silos and work towards a systemic approach that can be scaled nationally; we need nationally consistent data. But not in the way that we have traditionally approached the problem, where for each problem we request yet another view of council data, slightly different from the last possibly varying one or two dimensions. But almost always cast from the perspective of the requesting agency. Typically the capability of these data models to be repurposed or to evolve with changing requirements is very limited, see figure 2. In most cases as the requirements change the cost of repurposing previous work escalates, making it easier to start yet another data exchange table. In contrast we propose an object model developed from the perspective of Local Government. Founded upon the responsibilities and obligations managed by councils and cast from the perspective of managing services within the local community.

The approach is to build an integrated model that spans the services across the whole spectra of council. Starting with councils' core data sets the framework builds out in a modular approach. Each data set is effectively its own holon but when combined forms an integrated system.

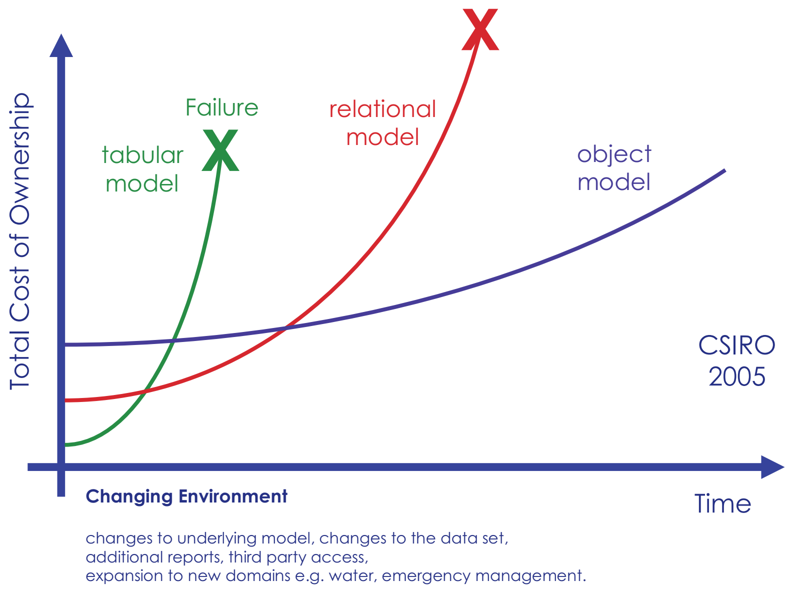


Figure 2. The Failure of Tabular Models to Support a Changing Environment. CSIRO 2005 unpublished.

The development of data modules is "Purpose Driven" where our fundamental purpose is

"the sustainable wellbeing of our communities"

The object of our concern may well be the physical asset, but the purpose that motivates the development of an information framework is not a bigger, better or perfectly preserved asset, rather a thriving and safe community.

Understanding that our purpose is larger than the asset itself is critical. The building of an information framework that serves the sustainable wellbeing of our community requires systemic thinking. It is this motivation that enables us to determine what information we need to manage and maintain. Furthermore, it is a motivation that is shared by other agencies [9].

All too often in public service the bar is set too low, the system is highly constrained within a very tightly defined terms of reference. Talk of a holistic or systemic approach is rhetoric; we do not expose ourselves to systemic thinking! We are incapable of thinking beyond our organisational boundaries. In asset management we have a situation where the organisational overlaps are multiple, the needs are different but not unrelated the motivation is fundamentally the same. We have an opportunity to do something different. To start thinking systemically, to break out of the silos and develop an information framework that serves multiple operational objectives.

The key to our success lies not in attempting to build some all encompassing system from the onset, but rather in defining an architectural framework that enables the progressive expansion of the core modules in an open and transparent manor. If our purpose is to support the sustainable well-being of our community then in order to make good decisions in planning for such a future the community and council require a holistic view of the municipality that provides an understanding of the interdependency between:

people and communities,

services,

resources – finance and depreciation,

physical environments,

governance and legal frameworks,

assets – infrastructure and amenities.

In order to help foster that understanding our asset management systems must support an information service that transforms the multiple, inconsistent and not so readily accessible asset data sources used to manage the diverse range of assets and the lifecycle of activities associated with them into a common view of assets that is harmonised with the other core services managed by council.

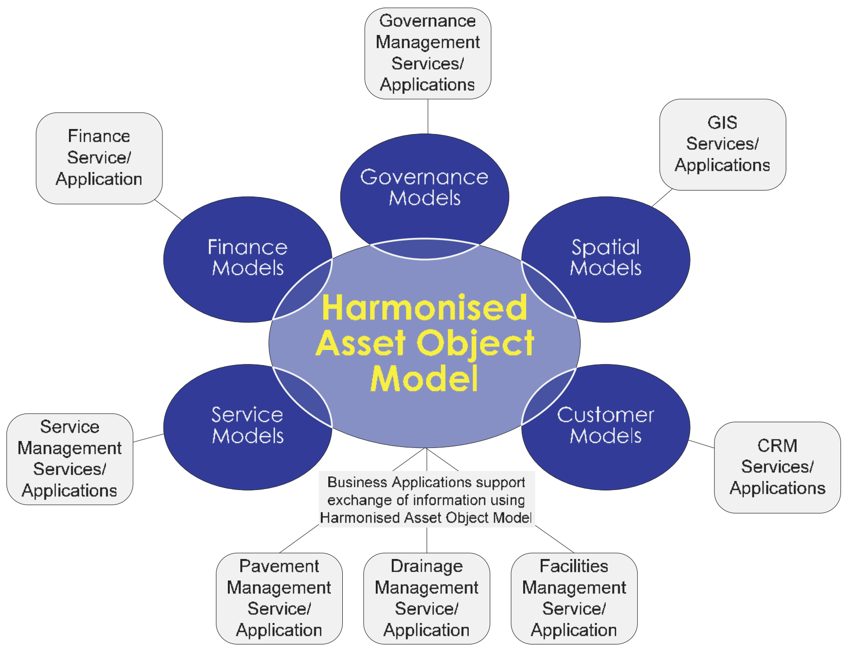


Figure 3. Harmonised Asset Reference Model Architecture

It is not just information about assets that our communities and council need; they require a whole range of knowledge. Because we are asset managers focused on the issues of renewal and the sustainability of our services the architecture diagram in Figure 3 deliberately represents the situation as an asset centric model. If we represented finance or customer management, we could transform the architecture to place these models at the centre. However, an asset centric starting point provides one of the best opportunities to act systemically.

The integrated information framework is not a solution to the Asset Renewal Gap, rather a necessary service that empowers people with information to make better decisions. The core component of the Information framework is a harmonised reference model. Essentially this is a common agreed reference model that enables the interoperable exchange and use of information relating to community assets.

The framework provides a mechanism for the exchange of information between the multitude of parties that have an ongoing need to access community asset data and the 600 local councils that are responsible for its collection and maintenance.

The Harmonised Reference Model is an object-oriented model of assets that councils manage on behalf of their community.

The Reference Model Contains:

1. Semantic Definition of the objects (people, roles, services, assets etc) that form part of the asset reference model.
2. UML static structure class model of objects.

Characteristics of the reference model are:

1. it is based on real world objects; as far as is possible the objects modelled are valid and readily recognisable constructs of the real world,
2. it has a modular component based structure, where objects are modelled once and may have multiple relationships,
3. it has the flexibility to manage inconsistencies; in building a harmonised model we recognises that there will be both a tight and loose fit between the model components and data stored in supporting business applications,
4. it is responsive to being viewed from multiple different perspectives; you should be able to ask lots of different questions of the same model,
5. there is a consistent interface used to access data across all components.

The methodology for developing the reference model is based on the object paradigm and endeavours to create as realistic a representation of the real world as possible. The outcome of this process is a model that remains robust and effective when viewed from multiple perspectives. The one model is better able to serve multiple purposes in supporting a diverse range of user needs.

8 THE BENEFITS

The benefits of an integrated information framework for asset data are that it will

1. enable the timely and efficient exchange of data,
2. ensure the interoperability of information exchanged across internal and external boundaries,
3. enable the aggregation of data from different areas,
4. facilitate the analysis and comparison of data.

The framework will lead to a more accurate and comprehensive understanding, at all levels of community - form local to national, of the services supported by our assets and the levels of investment required to sustain or improve the service outcomes.

The Information Exchange Framework will enable investment and funding organisations to substantiate wider planning decisions with factual data that is aggregated across organizational boundaries.

The agreed asset management ontology combined with the facility to exchange field values will deliver a more robust and better-informed platform for making decisions; in particular, where these decisions relate to the planning of sustainable development and services within our communities and the mitigation of hazards.

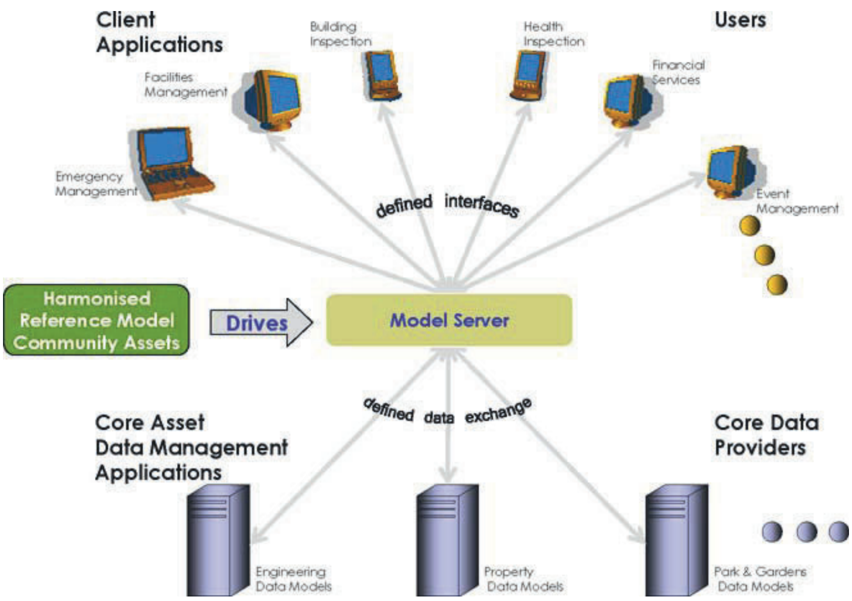


Figure 4. Integrated Information Framework Enables Multiple use of Asset Data

The end-to-end system will deliver more sustainable information management; the framework will provide the basis for a common standard for the representation of asset information. Thereby, reducing the requirement to support multiple exchange formats and bespoke system integrations. Data managers, such as utilities and councils will manage the data once, 3rd parties that require access to the data will obtain a consistent and agreed view of the data, minimizing the number of data mappings that have to be supported.

The information framework will enable councils and others to collect and maintain once, use many times without the repeated need to repurpose, see figure 4.

9 THE JOURNEY TO DATE

We set ourselves the following goals:

1. to develop an integrated information framework for community based asset management.

2. to establish technology environments for the interoperable exchange and management of asset data.

9.1 Progress

Our primary development has been in the area of Local Roads, working with both Queensland and Victorian Local Governments. We have developed a core object model based on the Queensland Roads Alliance data requirements. Work is currently being completed on a model driven server that will facilitate the interoperable exchange of physical data.

The model has as one of its foundations the concept of a network, which can be applied to other asset types such as drains, signs, footpaths etc. and work has just commenced on the extension of the model to include these assets.

Research Objectives

The outcomes of this project will be achieved through the following objectives:

The development of an "Integrated Information Framework" that will establish a methodology and ontology for the exchange of community based asset data. The framework will be instantiated through the development of reference models for local roads, drainage and irrigation networks and public furniture and signage.

The development of a harmonise interface with spatial modelling standards, namely, the ICSM Spatial Reference Model.

The development of a harmonise interface with Architectural and Civil Engineering (ACE) modelling standards, namely, the IAI IFC models.

The development of a model server to support the interoperable exchange of information between councils and other users, including shared service components and the extraction of data from backend systems.

10 NEXT STEPS

10.1 Framework Release

Over the next 12 months we intend to release further aspects of the framework, including:

1. Governance Arrangements regarding the knowledge and intellectual property generated in the Reference Model, including the process of acquiring consensus. The models are currently governed by an open source type license; the intent is to continue this form of license for the model but not necessarily the products that derive from it.
2. Partnership Arrangements and the mechanisms for dissemination of knowledge to practitioners and industry partners. There are many organisations operating in the area of public infrastructure and asset management. Our goal is to form strategic partnerships with key players to facilitate the acceptance, promotion and implementation of the model at a national and international level.
3. Publication of the Framework and Reference Model which represents the conceptualisation of our knowledge - an ontology; it will provide a controlled vocabulary that enables the description of relevant objects and the relations between them to be expressed in a formal way. It will provide a grammar for using the vocabulary terms to express meaningful concepts within asset management. The formal conceptualisation of our knowledge enables practitioners to make consistent, repeatable queries and assertions about our community assets.
4. Expansion of Scope of the network model to include water infrastructure assets.

10.2 Spatial Realisation

The ability to relate visually to models of the infrastructure/asset in planning and response is perceived by many practitioners as critical. Textual descriptions are considered insufficient to manage the complexities of the modern urban environment. There has been significant development in building similar reference models within the geospatial community. However, their focus has largely been on displaying the feature and its immediate attributes. The complex inter-relationships between assets their environment and the social context in which they exist has been ignored. Therefore, we need to bring these fields together in a harmonized interface.

Furthermore, as we continue to develop evermore-complex urban environments the need to exchange richer 3 & 4 dimensional information about the built environment becomes increasing relevant. The ACE community has been developing similar model to enable interoperability between all of the parties involved in building development. As a result the IAI has developed the Industry Foundation Class (IFCs) models for the interoperable exchange of 3D building data; which is now supported by all the major CAD vendors.

It is essential that we develop a harmonized approach with both the IAI and GIS communities, developing an asset reference model that is harmonized with these, whilst retaining its relevance in the broader community of users. In the next phase of work we will be developing a harmonise interface with spatial modelling standards, namely, the ICSM Spatial Reference Model and a harmonise interface with Architectural and Civil Engineering (ACE) modelling standards, namely, the IAI IFC models.

10.3 Model Driven Server

Without a tool to exchange information the reference models remain largely theoretical. A model driven server provides the realization of the benefits of the models to practitioners. In simple terms the model driven server is a data-store that can receive information from multiple sources, aggregate and store it in a managed environment that then allows other users to access that data; combining and manipulating it in ways that are relevant to their needs. The model is transparent and accessible to all, the model server provides the interoperability between systems.

Why model driven? Model driven primarily defines that the process of change is driven down from the model rather than up from a database implementation. The data-store is always maintained as a faithful representation of the reference model; change is driven by changing the model. A model driven server empowers practitioners to maintain control and influence over the reference model, whereas in more traditional approaches that control and influence is often usurped by the specialist technicians that maintain the database.

Queensland in partnership with the CSIRO are currently developing a model driven server for the Roads Alliance data, other similar servers are also planned but are currently not specified.

11 CONCLUSIONS

The asset renewal gap is a reality for Local Government. A solution to the gap requires debate at local, state and national levels. Its solution is not simply better asset management activities. What is needed is an honest dialogue with the communities about their aspirations for their communities.

If the dialogue is to be honest and meaningful participants must be informed.

Current information exchange is inadequate and difficult to aggregate at a national level. It is currently very difficult to present a consistent nation case for the renewal gap.

We propose that the problem is best solved by a national information framework that will enable the exchange of information between jurisdictions.

The framework should adopt a systemic approach to identifying and managing the necessary information.

The beneficiaries of such an approach would be the broad group of communities that all have a dependency on community assets. The collection and management of information would be handled once and used many times.

12 REFERENCES

[1] DoTaRS. (2003) Local Government National Report, Department of Transport and Regional Service, Canberra.

[2] DoTaRS. (2004) Local Government National Report, Department of Transport and Regional Service, Canberra

[3] Burns P. (2005) Review of Asset Management, AMQ International, Municipal Association of Victoria, Melbourne.

[4] Victorian Local Government Infrastructure Study. (December 1998) Facing the Renewal Challenge, Department of Infrastructure, Melbourne, 1998.

[5] MAV, (2006) The MAV Renewal Gap Model Program “A Statewide Approach”, Municipal Association of Victoria, Melbourne, 2006.

[6] MAV, (2006) The MAV Renewal Gap Model Program “A Statewide Approach”, Municipal Association of Victoria, Melbourne, 2006.

[7] Commonwealth Grants Commission. (2006) Draft report for the review of the interstate distribution of local road grants, Local Roads Conference, Canberra

[8] Commonwealth Grants Commission. (2005) Review of the Interstate Distribution of Local Road Grants Issues Paper CGC 2005/2, Canberra

[9] Department of Premier and Cabinet (2005) Growing Victoria Together, www.vic.gov.au

TECHNOLOGY MISSION ON INDIAN RAILWAYS

Nalinaksh S Vyas

Indian Institute of Technology, Kanpur 208 016, India

Abstract: Railways have been the instrument of technology development and growth in India. Introduction of railways in India is considered to have ushered in the industrial revolution and later facilitated industrial development. A Technology Mission on Indian Railways has been launched recently by the government. The effort is sponsored jointly by the Ministry of Human Resource Development, Ministry of Railways and a consortium of private industries. The goals and objectives of the mission include (i) development and adoption of state-of-art safety and control technologies through projects aimed towards achieving higher throughput, lower cost of transmission and safer train movement (ii) encouraging and initiating R & D activities pertinent to Railways in academic institutions and laboratories and establish convergence and synergy among them (iii) dissemination of technologies through participatory approach to other application areas. The paper describes some projects in technical detail and discusses issues pertaining to mission implementation and monitoring.

Key Words: Railway Technology, Indian Railways

1 INTRODUCTION

Indian Railways span a route of over 63,000 kilometers, over four different gauges. More than 13 million passengers travel on 14,000 trains and well over 1.4 million tones of freight are transported everyday in 6000 freight trains between 8000+ stations, aided by a fleet of 7,300 locomotives, 45,000 coaching vehicles and 280,000 goods wagons [1,2,3]. This massive network is in a process of continuous change and upgradation.

A Technology Mission has been recently launched to focus national attention and drive technologies in monitoring, control, communications, design, electronics and materials for Railway Safety. Earlier national programs on space and defence research did not merely achieve goals specific to the missions, but also provided impetus to technology endeavours in institutions all across the country. A Technology Mission on Railways will similarly help to initiate and incubate design and development projects of significant importance. The Mission objective is to develop and adopt state-of-the-art safety, control and design technologies defined by needs related to Indian conditions.

The projects currently in progress include Track Side Bogie Monitoring System; Derailment Detection Devices; Sensors for Detecting Hotboxes, Hot Wheels; On Board Diagnostics; Wheels and Axles of Improved Metallurgy; Corrosion Prevention Of Rails; Rail Flaw Detection Instrumentation; GPS based Rail Mapping & Train Navigation; Fog-Vision Instrumentation etc. Projects are executed through a trident consortium comprising of railway organisations, academic and research institutions and industry. The constituents of the consortium collaborate to bring domain specific expertise and share responsibilities. IIT Kanpur is the Nodal Agency. Features of various projects undertaken by the mission are described in this paper.

2 MISSION PROGRAMS

The Mission is divided into four Programs: each Program comprises of the following individual projects.

- Mission Program 1 (Traction and Rolling Stock): 1) Track Side Bogie Monitoring System 2) Derailment Detection Devices 3) Sensors for Detecting Hotboxes, Hot Wheels 4) On Board Diagnostics 5) Wheels and Axles of Improved Metallurgy 6) Measuring Wheel Technology
- Mission Program 2 (Tracks And Bridges): 1) Corrosion Prevention Of Rails 2) Improved Fastenings 3) Rail Flaw Detection Instrumentation 4) Environmental Friendly Coach Toilet Discharge System
- Mission Program 3 (Signals And Communications): 1) GPS based Rail Network Mapping and Train Tracking
- Mission Program 4 (Fog-Vision Instrumentation)
- The background and objectives of the individual projects are described below.

2.1 Trackside Bogie Monitoring System

The objectives of this project include (i) development of an automated system to be installed along the track for detecting faults in bogies of Rolling Stock and (ii) development of instrumentation for detection of falling components of the rolling stock which may cause derailment.

Reprofiling and replacing worn wheelsets constitutes a significant portion of rolling stock maintenance costs. Failure to detect misaligned wheelsets increases their removal rate. Misaligned wheelsets also responsible for increased fuel consumption and accelerated track deterioration. Visual inspections can detect only those wheelsets that have worn out appreciably. Typically, about 5% of wheelsets have abnormal behavior. The misaligned wheel sets represent a serious safety risk and earliest detection of misaligned wheelsets is essential.

Another concern is the falling parts of rolling stock. Occasionally, parts of rolling stock come off due to failure of fastening. These are normally detected by visual inspection. Falling parts may also cause derailment and there is need for automatic detection and warning

An optical system is being developed in this project to monitor (i) Angle of Attack (AOA), (ii) Tracking Position (TP), (iii) Inter-Axle Misalignment (IAM), (iv) Tracking Error (TE), (v) Angle of Attack (AOA). AOA of a wheel set is the angle between the track radial line and the center line of the wheelset's axle. For a good wheel set : AOA < 1 mrad.

TP of a wheel set is the position of the center of the wheel set with respect to the track centreline. IAM of a bogie is the angle formed between its two wheelsets (IAM = the angle of attack of the leading wheelset (AOAL) minus the angle of attack of the trailing wheelset (AOAT). TE of a bogie is lateral distance between the centers of the bogie's wheelsets (TE = TP of the leading wheelset (TPL) minus TP of the trailing wheelset (TPT)).

2.2 Derailment Detection Devices

This project envisages development of on-board equipment for sensing derailment possibilities of rolling stock. Development includes appropriate instrumentation and signal processing strategy and its integration with the existing brake system for minimizing losses due to dragging of derailed vehicle. Presently there is no instrumentation on Indian Railways for detecting derailment possibilities.

The process of derailment is characterized by heavy misalignment of the axle along with large oscillations and jerks. Vehicle dynamics software packages are being employed to carry out simulation of vehicles running on new or worn wheels. MEMS sensors for detecting vertical, horizontal accelerations and tilting have been identified and test runs are being conducted on Northern Railways. Recorded data is to be employed to arrive at suitable criteria for derailment detection.

2.3 Instrumentation for Hot Boxes and Hot Wheels

Hot wheels are caused by brake binding, while Hot Boxes are caused due to bearing failures. These may lead to rapid seizure, derailment, damage to rail, etc. Instrumentation is needed which can raise an alarm to driver or to the next station/control post through signalling or Radio Link, in case of development of abnormal conditions. The task involves development of an integrated approach to data acquisition and processing from various sensing systems – axle counter, trackside monitoring, vehicle identification, end of train identification, etc. The sensors need to detect box temperatures >80°C (or >25°C above ambient) during a very fast box transit time which is less than 4ms. In the case of Hot Wheels alarm should be raised on observation of temperatures of the order of 250°C, with similar sampling rate as for hot boxes.

2.4 Development of Wheels and Axles of Improved Metallurgy

Activities for developing steel of improved metallurgy for wheels and axles for better safety, reliability and enhanced carrying capacity without substantially altering design parameters. This is a large size project considering that Indian Railways currently have a population constituting of 1440000 wagon wheels, 280000 coach wheels, 72000 locomotive wheels and half the number of axles for each of the two; additionally several thousand new wheels are needed every year. The procedure adopted involves identification of some steel compositions and laboratory scale trials on variable solidification and thermo-mechanical processing parameters for the purpose of controlling the microstructures. Subsequently, their tensile properties, toughness, fatigue resistance and wear resistance are evaluated. The next step is production of steel samples on the pilot scale under optimized processing parameters. Data analysis and review is to be further carried out to standardize new processing parameters.

2.5 On Board Diagnostics

The objective of this project is to develop on-board diagnostics for diesel and electric locomotives through a microprocessor based control system. Diagnostics on existing locomotives is presently confined to trouble shooting knowledge of the driver. The present exercise includes development of appropriate instrumentation and signal processing strategy for various equipments which form part of the transmission and also for other auxiliary machines on board the locomotives. It will enable real time monitoring of vital locomotive equipments like prime mover, rotating machines, traction motor suspension bearings, axle bearings, radiator drive, compressor, transformer, tap changer, pantograph, etc on electric/diesel locomotives. The system will also have self-diagnostic features.

Typical problems in diesel locomotive and electric locomotive rotating machinery include Unbalance, Misalignment, Bending resonance, Oil Film Whirl, Cocked rotor, Shaft distortion, Mechanical looseness, Rotor rub, Gearbox defects, Asymmetric shaft etc. The diagnostic system will include on-line data-acquisition and display over multiple-channels simultaneously, frequency analysis and real-time FFT display, RMS value computation, trending analysis, data storage with date-time information, safe, tolerable and alarm limits for all channels, automatic visual and audio alarm in case of limit crossing, algorithmic diagnosis scheme and communication from rolling stock to the central control.

2.6 Measuring Wheel Technology

In this project instrumented wheel-sets are to be developed for measurement of the instantaneous vertical and horizontal forces at the rail-wheel contact point. Separate instrumented wheel-sets are being developed for locomotive, coach and wagons. Development of an Instrumented Wheel is important for a variety of reasons, particularly safety implications of rail-wheel interaction. The instrumented wheel set measures lateral and vertical forces and strains between a wheel flange and the rail-head. The relative magnitudes of the horizontal and vertical forces and their frequency characteristics hold the key to understanding of various wheel-track phenomenon. The Measuring Wheel is expected to serve as a major tool for carrying out studies on rail-wheel interactive forces, derailment analysis, wheel profile optimisation, bogie hunting etc. Measuring Wheel is also essential for analysis of various types of stresses coming on rail and wheels during rail wheel interaction before introduction of any new rolling stock in Indian Railways. The instrumentation includes strain-gauges and piezoelectric crystals which sense the dynamic phenomenon and transmit signals through telemetry. Equipment and the processors of primary handling data will be located on the axle of wheel set. Information from the rotating wheel pair will be transmitted in a digital code inside the coach for registration and processing.

2.7 Corrosion Prevention of Rails

This project investigates the problem corrosion, which is a major safety concern for railways. It affects the serviceable life of most track components, resulting in frequent replacement of rails. The major corrosion concerns are (i) the rail foot, under the liner location (ii) jamming of elastic rail clips (ERC) in the insert (iii) weldments and (iv) stray current corrosion. Corrosion begins at the rail foot which results into loss of rail section leading to fractures. Jamming makes it difficult to extract the leg of elastic rail clips (ERCs) from the insert for maintenance. Clips/inserts/sleepers get damaged during maintenance, while extracting the ERCs, resulting into heavy loss. The present work expects to specifically tackle corrosion under the liner location on rail foot and at weld collars.

2.8 Rail Fastenings

Rail fastening system consists of - Elastic Rail Clip or ERC, metal or GFN (glass filled nylon) liner, SG (Spheroidal Graphite) iron insert and grooved rubber pad. This project is aimed at stress and life analysis of the existing fastening system and its redesign.

2.9 Rail Flaw Detection Instrumentation

Common types of rail flaws include: transverse defect in rail head, Gauge face corner defect, rail weld defect (AT weld), bolt-hole defect, piping defect, half moon cracks at the weld. It is evident that flaw detection requires multiple probing activities. The objective is to build a non contact type ultrasonic testing system with sufficient storage and processing capabilities. The instrument vehicle will be operable manually at walking speed as well as speeds upto 80 km an hour.

2.10 Environmental Friendly Toilets

The stretch of the Indian railway network is long and in some cases a passenger may be required to spend up to three nights on a train. The existing coach toilet system in the Indian Railways' passenger trains consists of a lavatory in which the excreta are discharged directly to the ground through the lavatory chute. The design of the toilet is simple, robust, almost maintenance free and by and large convenient to a variety of users,. However, the present system raises major concerns regarding discharge of fecal matter on the track. These concerns include: damage to the rails, unacceptable aesthetic and hygienic/sanitary conditions, particularly on the railway stations, and non compliance to the environmental standards. Prototype physico-chemical and biological toilet systems are being currently investigated. Prototype residue management systems for installation in railway yards/stations are also being explored.

2.11 GPS Based Rail Network Mapping & Tracking

The objective of this project is to (i) develop an effective way to collect and disseminate information dynamically of every train in a given geographical boundary for its location, speed and direction of movement and (ii) ensure better and selective dissemination of information to passengers. Train tracking system using Global Positioning System (GPS) is being developed. Each train will have train locator unit to receive information from GPS satellites and continuously identify the position of train with information about train location (latitude and longitude values). GSM is to be used for connectivity and wherever needed use as an alternate location identifier. The data logger can also be used to provide services for a variety of applications like a central train enquiry system, anti- collision device, train charting etc.

2.12 Fog Vision Instrumentation

The project envisages development of instrumentation for improving the visibility during foggy weather. Train movement gets severely hampered during foggy climatic conditions in the northern part of the country. The foggy weather conditions consistently worsen during winter with fog getting more opaque. Such weather conditions prevail for nearly two months. Instrumentation is required to enable the train driver to see through the fog for uninterrupted and safe train operation. Information like position of obstacles on the track ahead, track condition etc. should be made available on a graphical console display. The visible distance should be equal to the normal visibility of the driver in day light conditions. Sensing technologies to be investigated Fog Vision System solutions considered are (i) mm wave Radar based imaging system (ii) mm wave Radiometer (iii) Infrared Radiometer (iv) Ultrasonic Sonar. Deploying multiple types of sensors may be essential to cater for different scenarios. In such cases, data from multiple sensors are to be fused intelligently to give a single display on the console.

3 MISSION IMPLEMENTATION ISSUES

Indian Institute of Technology (IIT) Kanpur and Research Standards & Development Organisation (RDSO) Lucknow, the R&D arm of Indian Railways are the major collaborators in the mission. A trident consortium comprising of (i) Academic and Research institutions, (ii) Railway Organisations and (ii) industry is formed for effective definition and implementation of projects. The constituents of the consortium collaborate to bring expertise and share responsibilities. RDSO brings extensive domain knowledge and experience in articulation of problems and conceptualizing activities. Academic institutions like IITs and CSIR laboratories contribute towards problem analysis, design synthesis and prototype development and the industry is expected to provide inputs relevant for adoption of technology and its commercialization.

The funding components are as follows: (i) Ministry of Human Resource Development, 50% (ii) Ministry of Railways 30% (iii) Industry 20%. The anticipated duration of the mission is 3 years.

A Mission Implementation and Coordination Committee (MICC) was constituted by the government to implement and coordinate various projects of the mission. The Director of IIT Kanpur is the Chairman of the Mission and the Director General of RDSO is its Co-Chair. The committee includes representatives from the Department of Science and Technology (DST), Government of India and the Council for Scientific and Industrial Research (CSIR). Additionally, industry is represented adequately in the committee. The major tasks of MICC include preparation of Mission Management Manual, project team constitution, budget allocation and periodic project monitoring. Industrial partners are suggested by the individual project teams. A memorandum of understanding is required to be signed between the academic/research institution, railways and the participating industry partner for each project before initiating activity.

Issues related to Intellectual Property Rights of various participating agencies are being presently worked out. However the agencies have been assured about adequate safeguard of their interests, in consonance with their financial share in the project. Funds are released to IIT Kanpur by the government and are centrally managed according to the advice of the MICC. Allocation of funds for individual project teams will be made on the recommendation of MICC. Each participating agency submits an audited Statement of Expenditure and Utilisation Certificate to IIT Kanpur at the end of each financial year.

For each project under the mission the following are included as essential components of deliverables: (i) study report on the literature study conducted for various causes of the problem; (ii) analysis results for the various causes, (iii) plan for validating the analysis results, (iv) report on validation of the analysis results with experimental data, (v) specification documents on wheel metallurgy, track and engine electronics, (vi) one set of prototype units with necessary hardware and software for the technology demonstration purpose, (vii) Manuals like User Guide, Quick Reference Guide, Trouble Shooting Guide for the hardware and the software delivered

Mass production and deployment is not included in the scope of the projects.

4 REMARKS

Railways are the backbone of India's economy and an integral part of the social fabric. Unlike developed countries Indian Railways operate under vastly challenging circumstances of overcrowding, low cost of travel, longer trains etc. The challenges posed in terms of technology development and implementation, are also thus greater, varied and unique in nature. These challenges also provide enormous opportunities for utilisation of available human and technical resources for development of safer and economical railway network.

5 REFERENCES

[1] Annual Statistical Statements, 2004-05, Indian Railways, The Directorate of Statistics and Economics, Ministry of Railways (Railway Board), Government of India, New Delhi.

[2] Roychoudhury S, (2005), The Great Indian Railway Atlas, Calcutta Art Studio, Kolkata.

[3] http://www.irfca.com

DEVELOPMENT OF THE RSM-BASED HYBRID EVOLUTIONARY ALGORITHM FOR LOW VIBRATION OF SHIP STRUCTURE

Young-Mo Kong[a], Su-Hyun Choi[a], Jin-Dae Songb, Bo-Suk Yang[b] and Byeong-Keun Choi[c]

[a]*Research Center, Daewoo Shipbuilding and Marine Engineering Ltd., 1 Aju-dong, Geoje, Gyungnam 656-714, Korea*

[b]*School of Mechanical Eng., Pukyong National University, San 100, Yongdang-dong, Nam-gu, Busan 608-739, Korea.*

[c]*School of Mechanical and Aerospace Eng., Gyeongsang National University, 445, Inpyung-dong, Tongyoung, Gyeongnam 650-160, Korea*

Abstract: This paper proposes a RSM-based hybrid evolutionary algorithm (RHEA) which combined the merits of the popular programs such as genetic algorithm (GA), tabu search method and response surface methodology (RSM). This algorithm, for improving the convergent speed that is thought to be the demerit of genetic algorithm, uses response surface methodology and simplex method. The mutation of GA offers random variety to finding the optimum solution. In this study, however, systematic variety can be secured through the use of tabu list. Efficiency of this method has been proven by applying traditional test functions and comparing the results to GA. It was also proved that the newly suggested algorithm is very effective to find the global optimum solution to minimize the weight for avoiding the resonance of fresh water tank that is placed in the after body area of ship. According to the study, GA's convergent speed in initial stages is improved by using RSM method. An optimized solution is calculated without the evaluation of additional actual objective function. In a summary, it is concluded that RHEA is a very powerful global optimization algorithm from the view point of convergent speed and global search ability.

Key Words: Genetic Algorithm, Response Surface Methodology, Tabu Search Method, Tabu List, Simplex Method

1 INTRODUCTION

Many dynamic analyses are focused on finding the maximum response and avoiding the resonance in a given structure under all excitation forces [1, 2]. Usually, these features provide the basis of a design limit and are thus employed to determine the dynamic characteristics of a structure and its weight. For this reason, weight minimization for reducing the response and avoiding resonance has always been a major concern to vibration analysts. Many classic optimization methods and commercial softwares have been developed and most of them are very effective, especially to solve the local optimum problems. However they are not so good tools to find a global optimum solution for the system. To overcome this disadvantage, many search algorithms have been developed seeking for a global optimum solution. One of the most popular methods is the genetic algorithm (GA) [3, 4]. The GA is an optimization algorithm loosely inspired by evolutionary and is a powerful and general global optimization method. It does not require the strict continuity of classical search techniques, but it allows non-linearity and discontinuity to appear in the solution space. Due to the evolutionary characteristics, the GA can handle all kinds of objective functions and constraints defined on discrete, continuous, or mixed search spaces. However, the global access of the GA requires a computationally random search. So, the convergent speed to the exact solution is slow. Furthermore, the coding of the chromosome for a large dimensional problem will be very long, in order to get a more accurate solution. This results in a large search space and huge memory requirements for the computation. To overcome these demerits, many researchers have studied to develop many hybrid genetic algorithms which combine genetic algorithm with other ones [5, 6]. These can save computation time and find the global solution as far as it goes. However, new algorithms are required for better accuracy and faster convergent speed to get an optimum solution in the complicated and big structures like ships.

To seek for the optimum solution of multi-peak function with high accuracy and high speed, a new hybrid evolutionary algorithm is suggested in this study. This combines the merits of the popular programs such as GA, tabu search method, response surface methodology (RSM) and simplex method. The algorithm, for improving the convergent speed that is thought to be the demerit of GA, uses RSM and simplex method. The mutation of GA offers random variety to finding the optimum solution. However, systematic variety can be secured through the use of tabu list of tabu search method. Especially, in the initial stages, GA's convergent speed can be improved by using RSM which is using the information on the objective function acquired through GA process and then making response surface (approximate function) and optimizing this. The optimum solution is calculated without the evaluation of an additional actual objective function, and the GA’s convergent speed is improved. Efficiency of this method has been proven by applying traditional test functions and comparing the results to GA. It was also proved that the newly suggested algorithm is very effective to find the global optimum solution to minimize the weight change for avoiding the resonance of fresh water tank that is placed in the after body area of ship.

2 CONCEPT OF RSM-BASED HYBRID EVOLUTIONARY ALGORITHM (RHEA)

The main idea is to reduce the evaluation number of the objective function by using RSM which is one among the designed experiments to reduce the repetitive number, since it is one of the demerits of optimum design. The RHEA consists of four main categories: GA for governing the general algorithm; tabu-list for systematic variety of solution; RSM for improving convergent speed for getting a candidate solution; modified simplex method for local search. Fig. 1 represents the flowchart of the RHEA. The left side of the flow chart shows global search region that is similar to the flowchart of standard genetic algorithm, excluding the function assurance criterion (FAC), Sh (part A), tabu-list (part B), and RSM (part C). These parts offer candidate solutions, which are considered as initial search points in the local search region. The right side represents the local search region. This part finds out the optimum solution by the modified simplex method, which use the final solution by results of global search as initial search point.

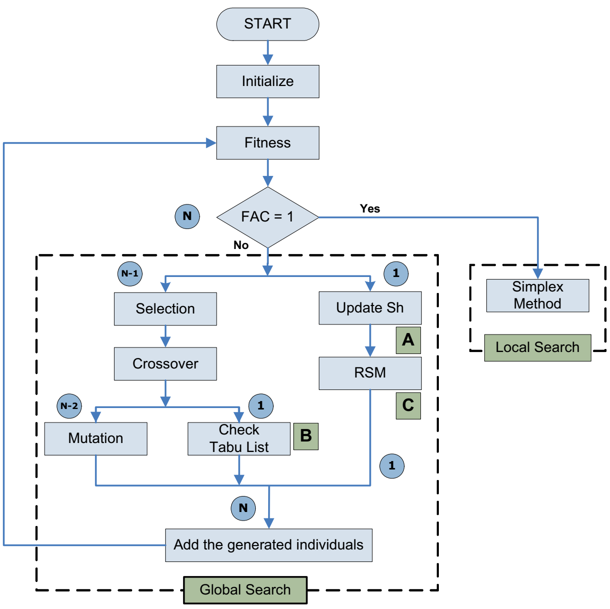


Fig. 1 Flowchart of RHEA

Part A in Fig. 1 shows the Sh region which provides the well distributed points to make a response surface. Part B in Fig. 1 shows that the tabu list is checked to have a diversity of solution. The one individual which is selected in GA's individuals after crossover is reviewed to secure the diversity of solution. If diversity of solution is secured, we select the individual and if not, we repeat the crossover process. That is, individual is selected when it is located far away from the dense area. Part C in Fig. 1 represents an RSM region. It is largely divided by 3 parts. Firstly, considering the boundary condition in the response surface for optimization, the upper and lower values of design variables can be considered in this calculation process. However, the merits of this method are diminished when addition constraints like natural frequency are considered, because it has to evaluate the objective function to get the results from external calculations. To overcome this problem, this study used Sh as training data and inferred the satisfaction of constraint condition using RBF network [7]. In this way, calculation of actual problems could be avoided. Secondly, it makes a response surface from Sh by using the least square method (LSM). Finally, the optimum solution of the response surface is calculated by using tabu search method. To increase optimization speed, gradient based algorithm can be used. However, the solutions satisfying constraint condition cannot be guaranteed since the constraint condition is difficult to define precisely. Also we adopt tabu search method which has an excellent initial convergent speed, because the implementation of the response surface concept is to search for the approximate candidate solution.

3 PROCEDURES OF RSM-BASED HYBRID EVOLUTIONARY ALGORITHM (RHEA)

RSM-based hybrid evolutionary algorithm (RHEA) is introduced as follows:

Step 1: Generate the initial chromosome $v_k\ (k = 1, 2, \cdots, pop\_size)$ randomly with n elements. *Step 2*: Generate the initial solutions, and estimate constraint and set up a parameter range. *Step 3*: Evaluate the fitness of individuals. *Step 4*: Evaluate the FAC, if it is satisfied, go to step 12 otherwise go to step 5. *Step 5*: Update Sh(a set of history) : $Sh = \{(X_{sh}, F) | X_{sh}\ R^N, F\ R\}$. where $X_{sh} = [x_1, x_2, \cdots, x_i, \cdots x_N]$, N is number of design variables. *Step 6*: Selection. *Step 7*: Crossover and check tabu list. *Step 8*: Construct response surface (RS) from Sh.

$$f_{rs} = \alpha_0 + \sum_{i=1}^{N} \alpha_{ii} x_i + \sum_{i=1}^{N} \alpha_{ii} x_i^2 + \sum_{i=2}^{N} \sum_{j=1}^{i-1} \alpha_{ij} x_i x_j \qquad (1)$$

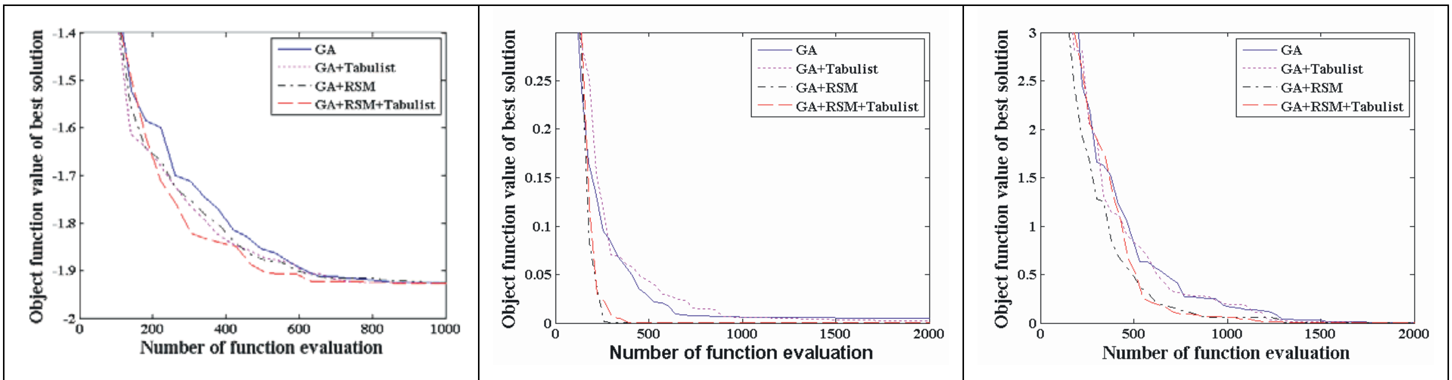
where α_0, α_{ii}, α_{ij} are coefficients calculated by LSM.

Step 9: Train RBF network by Sh to construct the constraint conditions approximately. *Step 10*: Calculate the optimum design on the response surface by tabu search method and generate one individual based on X*. *Step 11*: Mutate and go to step 3. *Step 12*: Search the optimum solutions by the local concentration search (modified simplex method) for best candidate.

4 NUMERICAL EXAMPLES OF SEVERAL FUNCTION OPTIMIZATIONS

Three test functions are used to verify the efficiency of the proposed hybrid algorithm: the first one is the four-peak function [8], which has one global optimum with three local optima; and the second one is Rosenberk's function [9] which is known as banana function and has just one global optimum; and the last one is the Rastrigin function [10] which has one global minimum with 220 local minima.

Fig. 2 represents the convergent trend of objective function for each test function. According to the results, GRSM (GA+RSM) and GRSMT (GA+RSM+Tabu list) algorithms which are based on RSM have faster convergent speed and more accurate solutions than GA, which validated the efficiency of RSM on the calculation. Also tabu list enables convergence to solutions quickly on the multi-peak function due to the systematic diversity of solution. Table 1 shows the comparison of optimization results for the above stated three test functions. The evaluation number means total evaluation number of the objective function used in optimization procedure, and it is directly proportion to the total calculation time. According to the results, for all test functions, RHEA can give better solutions than GA on accuracy and convergent speed. For the Rastrigin function, which is very useful to evaluate the global search ability because there are many local minima around the global minimum, RHEA found global minimum with higher accuracy and less elapsed time compared to GA. According to these results, the proposed new hybrid algorithm is a powerful global optimization algorithm from the view of convergent speed and global search ability.



(a) Four-peak function | **(b) Rosenbrook function** | **(c) Rastrigin function**

Fig. 2 Convergent trend of objective function

Table 1

Comparison of optimization results

Test function	Exact solutions	Methods	Results $f(x)$	(x_1, x_2)	No. of evaluation
Four-peak	$f(x) = 1.9543$	GA	1.927	2.403 e-3, 2.787 e-3	2353
function	$x_1 = x_2 = 0$	RHEA	1.927	2.736 e-3, 2.736 e-3	459
Banana	$f(x) = 0$	GA	1.640 e-5	9.960 e-1, 9.960 e-1	1046
function	$x_1 = x_2 = 1$	RHEA	0.0	1.0, 1.0	419
Rastrigin	$f(x) = 0$	GA	1.586 e-4	1.408 e-4, 8.15 e-4	2109
function	$x_1 = x_2 = 0$	RHEA	0.0	-3.076 e-9, -7.747 e-10	514

5 APPLICATION FOR OPTIMUM DESIGN OF FRESH WATER TANK OF SHIP

In the engine room and the after body area of the ship, there are so many tank structures in contact with fresh water, sea water or oil. Also these possibly subject to the excessive vibration during voyage because they are arranged around the main excitation sources of ship such as the main engine and propeller. If problems occur, it takes a lot of cost, time and effort to improve the situation because the reinforcement work for empting the fluid out of the tanks, additional welding and special painting and so on is required. It is therefore very important to predict the precise vibration characteristics of the tank structures at the design stage. Optimum design needs to be applied. Especially when the structure is in contact with fluid much analysis time is taken. So, a new hybrid optimization algorithm is required for getting a short analysis time and accurate solution. In this study, optimum design of a fresh water tank in an actual ship is carried out to verify the validity of the proposed optimization algorithm (RHEA) and the results are compared to that of GA.

5.1 Vibration Analysis of Fresh Water Tank

It is difficult to predict the vibration response of a local structure due to the complicated transfer mechanism of excitation force and the difficulty of assuming the damping ratio. Traditionally, therefore, the vibration analysis considering the design of avoiding resonance is conducted to prevent the local vibration.

In this study, the vibration analysis of the fresh water tank is carried out using NASTRAN which is a commercial finite element program and widely used for big structures like ships. Fig. 3 shows the model and arrangement of the fresh water tank. Fig. 4 shows the design variables and boundary condition of the fresh water tank. Considering the precision of analysis and time consuming modeling process, the range of modeling of fresh water tank is constrained to one side of the tank. The boundary conditions for the model are specified: the simple supports are used to the tank boundary area which is connected to the other bulkhead and deck.

In general, the design for avoiding local structure resonance in ships requires that the natural frequency of the structure must be two times higher than the blade passing frequency of the propeller under the maximum rpm of the main engine. In this study, design target frequency is set as above 14.02 Hz which considers safety margins and twice blade passing frequency of the propeller (12.13Hz).

Fig. 5 shows the first three modes and natural frequencies of the fresh water tank by NASTRAN. These three modes frequently occurred on the fresh water tank during voyage. Especially, the 1st mode (8.60 Hz) is a stiffener (stringer) mode which generates a strong vibration and much effect on the structure. In this model, the 1st natural frequency of the structure is also within the resonance region where twice blade passing frequency of propeller is 12.13 Hz. Therefore, the natural frequency of structure is needed to be increased up to the target frequency under the condition that the tank is fully filled. The natural frequency of structure which is contacting with fluid can be changed according to the water line of the tank. So, in order to design a safe structure, the three modes of the fresh water tank are concerned in this study.

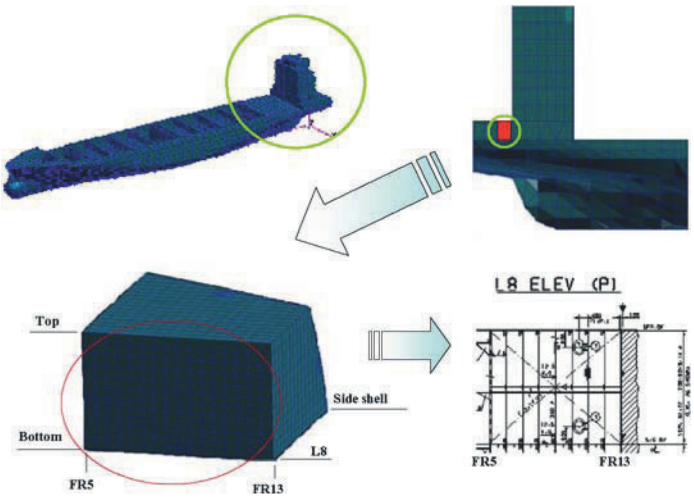
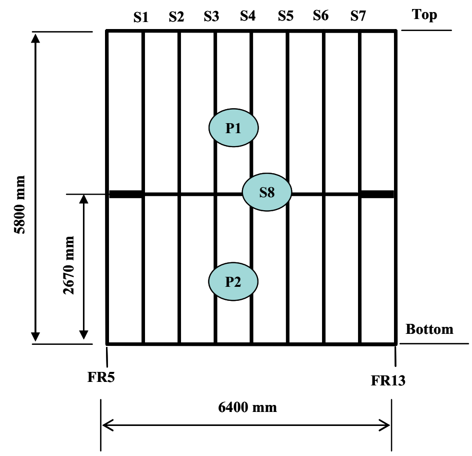
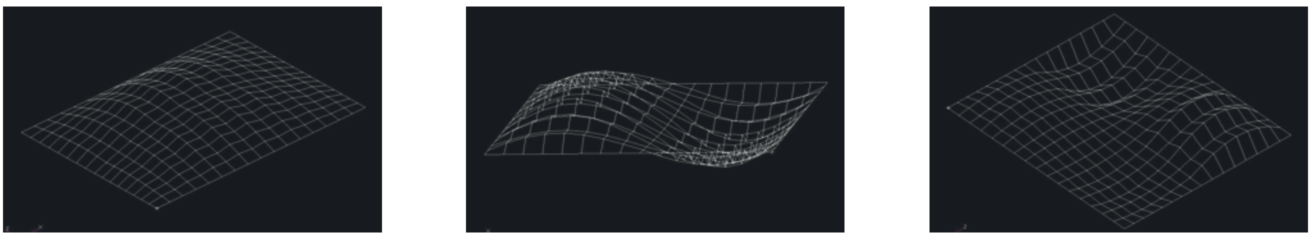


Fig. 3 Model and Arrangement of fresh water tank



S1	Stiffener 1
S2	Stiffener 2
S3	Stiffener 3
S4	Stiffener 4
S5	Stiffener 5
S6	Stiffener 6
S7	Stiffener 7
S8	Stringer 1
P1	Plate 1
P2	Plate 2

Fig. 4 Design variables and boundary conditions of fresh water tank



(a) 1st mode (8.60 Hz) **(b) 2nd mode (18.82 Hz)** **(c) 3rd mode (19.17 Hz)**

Fig. 5 Mode shapes of fresh water tank

5.2 Optimum Design of Fresh Water Tank

Design variables: The main vibration modes on the fresh water tank are stiffener modes in transverse direction. One of the most important factors is the stiffness of stiffeners. In this study, the stiffener size and plate thickness of fresh water tank in Fig. 4 are defined as design variables in Eq. (2).

$$x = \{S1\ S2\ S3\ S4\ S5\ S6\ S7\ S8\ P1\ P2\}^{T} \quad (2)$$

where S and P mean stiffener size and plate thickness, respectively.

Constraints: The web length of stiffener L_w is restricted as two categories like Eq. (3) according to the shipyard's practice.

$$150 \leq L_w \leq 450 \text{ mm for stiffeners } (S1\text{-}S7), \quad 500 \leq L_w \leq 1000 \text{ mm for stringer } (S8) \quad (3)$$

The basic concept of local vibration design is also the minimization of the response at each point. However, it is difficult to evaluate how much the excitation force influences on local structure. So, in this study, natural frequency of the structure is restricted as Eq. (4) which considers a safety margin of twice blade passing frequency of the propeller.

$$\omega_n \geq 14.02 \text{ Hz} \quad (4)$$

Objective function: The objective function combines linearly the weight of fresh water tank with natural frequency of structure like Eq. (5). The objective is to get an economic and sound structure to reduce the weight of stiffener and to increase the natural frequency.

$$\text{Minimize } f(x) = \alpha\left(\frac{W_1}{W_0}\right) + \beta\left(\frac{\omega_t}{\omega_0}\right) \quad (5)$$

where, W_1 and W_0 mean current and original weight of stiffeners and plates, respectively. ω_t and ω_0 mean target and current natural frequency, respectively. α and β are weighting factors ($\alpha = 0.5$, $\beta = 0.5$).

5.3 Optimization Results and Discussion

The optimum design was carried out to get an optimal size of stiffener and plate thickness on the fresh water tank to maintain the anti-vibration design of it. Table 2 shows the results of the design variables before and after optimization. It shows that the stringer S8 is increased by 72% and the others by 4.0-52%. This result indicates that the most reasonable modification method is to increase the stringer which has an effect on the decreasing the span of the vertical stiffeners. In this case, however, the plate thickness does not have any effect on the natural frequency of the structure. Table 3 shows the variation of natural frequency and weight of structure before and after optimization. According to the results, the 1st natural frequency increased by 163 % from 8.6Hz to 14.02Hz, and the safety margin with twice passing frequency of the propeller correspondingly changed from -29.1% to 11.56%. Therefore, the structure is free from resonance. Moreover, the weights of stiffeners which are applied to the design variables also decreased in spite of higher natural frequency. In summary, the local vibration problems which require avoidance of structure resonance through the movement of natural frequency without additional weight has been successfully solved by the proposed optimization method. Table 4 and Fig. 5 show the comparison of optimization results between GA and RHEA. The evaluation number means a total evaluation number of the objective function used in the optimization procedure, and is directly proportional to the total calculation time. According to the results, RHEA can give better solutions than GA on accuracy and convergent speed. These results lead us to draw the conclusion that the proposed new hybrid algorithm is a more powerful global optimization algorithm from the view of convergent speed and global search ability.

Table 2

Comparison of original and optimal design variable

Design variable		S1	S2	S3	S4	S5	S6	S7	S8	P1	P2
Original		200	200	200	200	200	200	200	550	11.0	11.0
Optimum	GA	214	320	253	325	328	277	281	893	10.7	10.6
	RHEA	207	223	285	283	303	251	230	947	10.3	10.0
Remarks (%, RHEA)		4.0	12.0	43.0	42.0	52.0	26.0	15.0	72.0	-6.36	-9.09

Table 3

Comparison of results

Item	Original	Optimum	Remarks
Natural frequency	8.60 Hz	14.02 Hz	163 %
Weight	4883kg	4652 kg	-4.73 %

Table 4

Comparison of optimization results

Item	Weight	Objective function	No. of evaluation
GA	5001 kg	0.5547	1846
RHEA	4652 kg	0.5167	1638

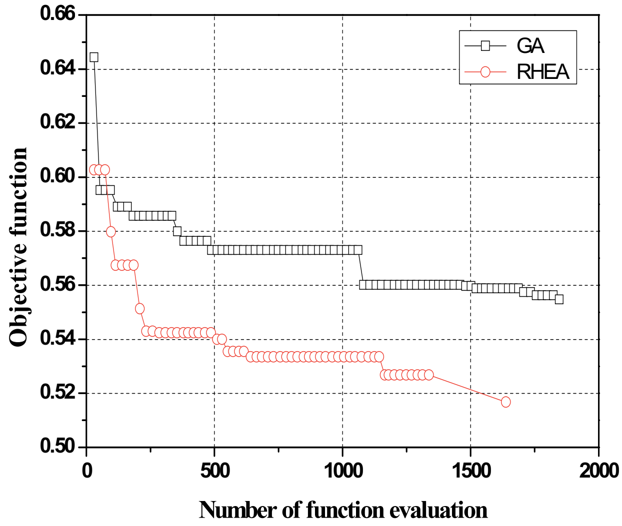


Fig. 5 Convergence of objective function

6 CONCLUSIONS

This paper introduces RSM-based hybrid evolutionary algorithm, as a new kind of a hybrid optimization algorithm that combines the merits of the popular programs such as genetic algorithm, tabu search method and response surface methodology. This algorithm, for improving the convergent speed that is thought to be the demerit of genetic algorithm, uses response surface methodology and simplex method. The mutation of GA offers random variety to finding the optimum solution through the use of tabu list. Especially, in initial stages, GA's convergent speed can be improved by using RSM which is using the information on objective function acquired through GA process and then making response surface (approximate function) and optimizing this. The optimum solution is calculated without the evaluation of additional actual objective function, and the GA's convergent speed is improved. Efficiency of this method has been proven by applying traditional test functions and comparing the results to GA. It was also proved that the newly suggested algorithm is very effective to find the global optimum solution to minimize the weight for avoiding the resonance of fresh water tank that is placed in the after body area of ship. Finally it is concluded that the proposed new hybrid algorithm (RHEA) is a very powerful global optimization algorithm from the view point of convergent speed and global search ability.

7 REFERENCES

[1] Kitamura M, Nobukawa H and Yang F (2000) Application of a genetic algorithm to the optimal structural design of a ship's engine room taking dynamic constraints into consideration, Marine Science and Technology, 5, pp. 131-146.

[2] Yang BS, Kong YM, Choi SH, Chae SI, Song JD and Kim YH (2005) Development of NASTRAN-based Optimization Framework for Vibration Optimum Design of Ship Structure. Trans. of the Korean Society for Noise and Vibration Engineering, 15(11), pp.1223-1231.

[3] Goldberg DE. (1989) Genetic Algorithms in Search, Optimization & Machine Learning, Addison-Wesley Publishing Company, pp. 1-146.

[4] Davis L (1991) *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York, pp. 3-23.

[5] Kim YC and Yang BS (2002) An enhanced genetic algorithm for global and local optimization search. Trans. KSME (A) 26(6), 1008-1015. (in Korean)

[6] Sato T and Hagiwara M (1998) Bee system: finding solution by a concentrated search. T. IEE Japan, 118-C(5), 721-726.

[7] Howard D, Mark B and Martin H (2005) *Neural Network Toolbox for Use with MATLAB*. The MathWorks.

[8] Kim YC (2003) *Development of Enhanced Genetic Algorithm and Its Applications to Optimum Design of Rotating Machinery*, Ph.D. Dissertation, Pukyong National University, South Korea

[9] Homaifar A., Qi C and Lai S (1994) Constrained optimization via genetic algorithm simulation. Electronics Letter, 62(4), 242-254.

[10] Kim YH, Yang BS, Kim YC and Lee SJ (2003) Bearing parameters identification using hybrid optimization algorithm. in Proceedings of 32nd International Congress and Exposition on Noise Control Engineering, Jeju, Korea, pp. 4212-4219.

ENTERPRISE RELIABILITY MANAGEMENT

Jason Apps

ARMS Reliability Engineers

Abstract: Traditionally reliability effort and Reliability Centered Maintenance (RCM) studies have been conducted on a project basis and one of the most common pitfalls to a successful outcome is the effort involved in implementation of the resulting maintenance strategies. It is all too common for RCM plans to be developed which result in a significant reduction of business risk only to be never completely implemented. This paper outlines an enterprise approach to reliability analysis allowing for an efficient implementation or RCM projects. The benefit of the approach also allows for benchmarking across an organisation, performance tracking, triggering of analysis projects and efficient sharing of analysis data and results.

Key Words: Reliability Centred Maintenance, RCM, Analysis, SAP

1 INTRODUCTION

Traditionally reliability effort and Reliability Centered Maintenance (RCM) studies have been conducted on a project basis and one of the most common pitfalls to a successful outcome is the effort involved in implementation of the resulting maintenance strategies.

It is all too common for RCM plans to be developed which result in a significant reduction of business risk only to be never completely implemented.

This paper outlines an enterprise approach to reliability analysis allowing for an efficient implementation or RCM projects. The benefit of the approach also allows for benchmarking across an organisation, performance tracking, triggering of analysis projects and efficient sharing of analysis data and results.

2 IMPLEMENTATION

The outputs from an RCM study are the individual tasks which combine according to set business rules to form the maintenance plans. The implementation process generally involves either manual upload of SAP tables or an import using the Legacy System Migration Workbench (LSMW). The LSMW is capable of uploading data from excel spreadsheets into SAP tables provided the excel spreadsheets are configured correctly.

It is essential to have a resource familiar with the SAP tables to be able to correctly create all the upload tables required, due to the complex nature of the SAP table structure. If the LSMW is to be used the user also needs experience in the correct use of LSMW.

The tasks are combined into “Task Groups” or Maintenance Plans as per the business rules and work instruction type documents are completed ready for upload and linking.

Due to the nature of the resources involved to complete the process it involves interfaces between several departments and at times systems to complete the implementation. This factor alone greatly reduces the likelihood of an efficient implementation.

2.1 The living System

With rapidly changing business environments, and with most plants changing operating conditions, demands and procedures the maintenance plans require periodic review.

Most RCM analyses are completed within tailored computer programs which, depending on the chosen tool, allow relatively rapid update of parameters and therefore review of the maintenance plans. The challenge arises in updating EAM data to reflect any interval and maintenance plan detail changes.

For example, for a particular piece of equipment there may be a 3 monthly and a 6 monthly strategy, each of which contain a number of tasks. During a review it may be that one task which was conducted 6 monthly now has a recommendation to change to 3 monthly. In this case someone has to update several relevant SAP tables and also manually locate the appropriate work instruction documents, manipulate them and then reload to the correct location.

2.2 Interfacing Options

Essentially there are three options available to transfer data between the reliability project software and the EAM system.

1 Manual export and data update

2 Using standard data import applications (i.e. LSMW for SAP)

3 Enterprise Integration

2.3 Manual export and data update

The simplest method of data transfer is to actually export data from either system and then manually key information in to the source system. For example work order history can be extracted into MSExcel spreadsheets, analysed and then either manually keyed or using Windows Cut and Paste feature to enter data into the reliability project software.

By the same token resulting maintenance plans can be exported from the reliability project into a MSExcel spreadsheet and then from these exports data can be consolidated and segregated such that it can be manually keyed into the EAM system.

Advantages	Disadvantages
- Requires little pre-thought or planning with regard to consistency	- Difficult to work with large data volumes - Extremely time consuming data input and manipulation - Difficult to assure standards or consistency of analysis or ERP configuration - Extremely difficult to make large scale changes to existing SAP objects - High level of skills and knowledge required to extract, manipulate and upload data - Association between the maintenance strategy and its analysis is lost after upload to SAP - Low sustainability rate, i.e. overwhelming nature of human intervention and susceptibility to error leads to questioning of value and loss of motivation

2.4 Standard Data Import Applications

This method improves the efficiency of the manual process by utilizing standard data import applications which will read an MSExcel spreadsheet and import into the EAM data tables.

For example users of SAP can use the Legacy System Migration Workbench (LSMW) application to import a MSExcel spreadsheet. The spreadsheet needs to be configured correctly prior to the upload requiring the user to have a sound knowledge of the SAP data tables and their particular configuration for that installation.

Advantages	Disadvantages
- Allows faster methods of data entry compared with manual methods - Limited reuse of template files and configurations	- Difficult to work with large data volumes - Time consuming data input and manipulation - Difficult to assure standards or consistency of analysis or ERP configuration - Extremely difficult to make large scale changes to existing SAP objects - High level of skills and knowledge required to extract, manipulate and upload data - Association between the maintenance strategy and its analysis is lost after upload to SAP - Low sustainability rate, i.e. overwhelming nature of human intervention and susceptibility to error leads to questioning of value and loss of motivation

2.5 Enterprise Approach

Taking an enterprise approach provides a significant improvement in efficiency, accuracy and functionality by providing a robust interface along with methods of automatically configuring data in the required format.

Advantages	Disadvantages
– Extremely fast access to ERP Data – Object identity is maintained so much easier to find in the ERP system – Fast and easy uploading to ERP System – Assurance of ERP object and data standards – Works with large data volumes – Can make changes to large volumes of data – Association between the maintenance strategy and its analysis is retained after upload to SAP – High sustainability rate due to automated nature of solution and visibility of the ongoing value delivered	– Requires planning and pre-thought with regard to technical object design – Initial configuration effort

2.6 Comparison of Options

The following table outlines the functionality comparison of the interfacing options.

Feature	Manual	Spreadsheet	Integrated
Project capability	✓	✓	✓
Reliability block modeling	✓	✓	✓
Equipment tactics	✓	✓	✓
Life cycle costing	✓	✓	✓
Initial optimization	✓	✓	✓
Ongoing optimization	X	X	✓
Multi-project capability	X	X	✓
Enterprise capability	X	X	✓
Scalable	X	X	✓
Library management	X	X	✓
Closed-loop management	X	X	✓
Compliance alerts	X	X	✓
Predicted/actual comparison	X	X	✓
Portal publication	X	X	✓
KPIs	X	X	✓
Systematic SAP connectivity	X	X	✓
Monitor application of Corporate standards	X	X	✓

The Enterprise approach yields significantly greater functionality, whilst providing a sound basis for organisations to continually improve their maintenance strategies over time and also maximize the leverage from any reliability project completed by allowing benchmarking and templating across and entire organisation.

3 ENTERPRISE RELIABILITY MANAGER

The Enterprise Reliability Manager (ERM) can dramatically increase the efficiency of implementation by automating the grouping of tasks, creation of appropriate SAP tables and generation of work instruction templates. ERM can update SAP tables and removes the need to use another import facility such as LSMW.

ERM uses defined business rules to generate all tables and data and so ensures a consistent approach to data structure. ERM also conducts data integrity checks to ensure data tables are valid prior to updating in SAP.

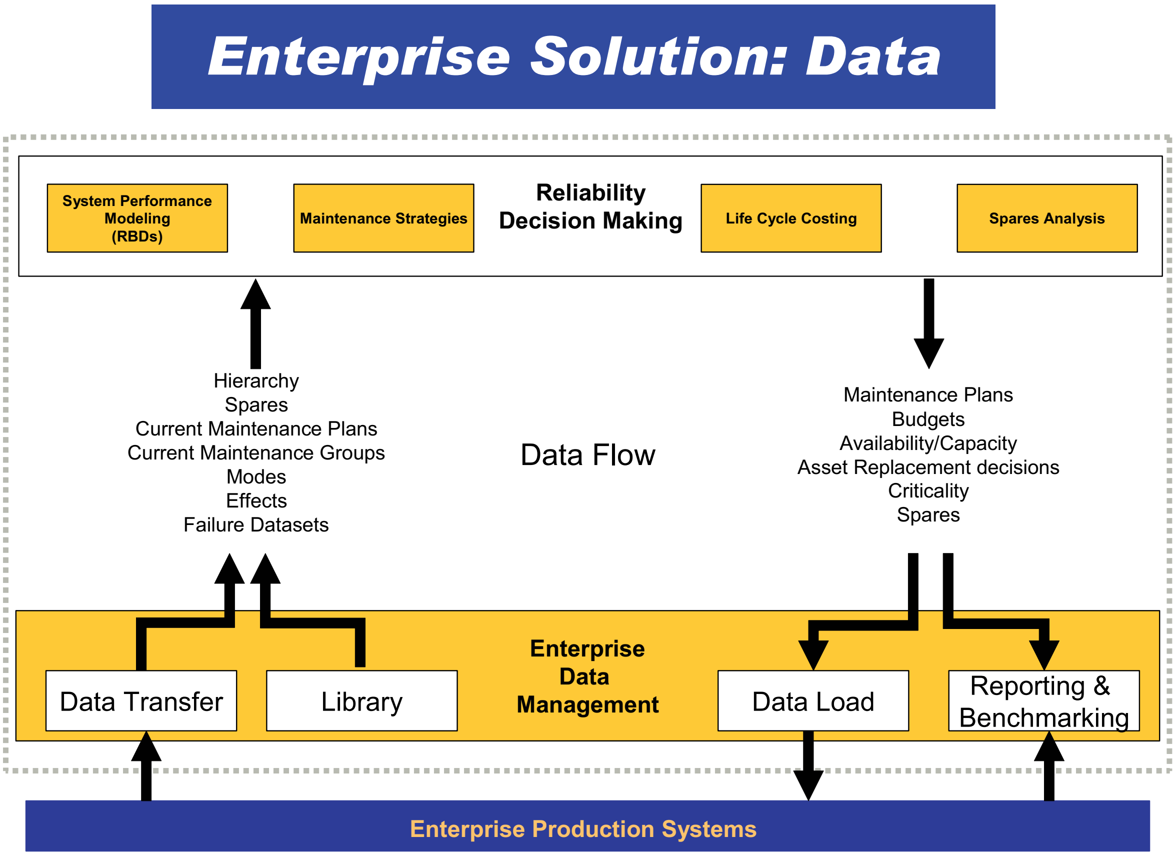


Figure 1 – Data flow within an enterprise level solution

The addition of an enterprise level solution provides a vehicle to take a truly site or organisation wide approach to reliability management. Since the ERM can upload/download data from SAP, a single instance of data can be maintained. Any reliability modeling or RCM analysis can be based on actual data from SAP, and the results can be stored for use as a template for the analysis of similar assets.

The ERM can then be used to provide benchmarking data across sites, comparisons of analysis for similar equipment and KPI reporting by tracking actuals against any predictions. The ERM monitors actual performance metrics and can be set to flag any deviations for analysis or review.

The ability to track the success of implementation is of great importance in terms of tracking the success of a reliability focus whilst also allowing a focus on continuous improvement and update by identifying deviations from predictions.

3.1 Enterprise Reliability Manager Applications

An enterprise approach to interfacing EAM data with reliability projects has three distinct categories which are considered below.

Addition of Assets – Unique assets

The category represents the case where a maintenance strategy needs to be developed for an asset and there are no similar assets within the organisation with useful information to use as a baseline.

In this category there may not even be an asset hierarchy in which case the reliability project provides the basis for the asset hierarchy, development of maintenance strategies and task lists.

Additional of Assets – Similar Assets

This category represents the case where a maintenance strategy is required for an asset which is similar to other assets within the organisation.

In this case an existing asset hierarchy and maintenance plans are utilized as a basis of the reliability project. On completion of the project an asset hierarchy, maintenance strategy and task list will be loaded into the EAM.

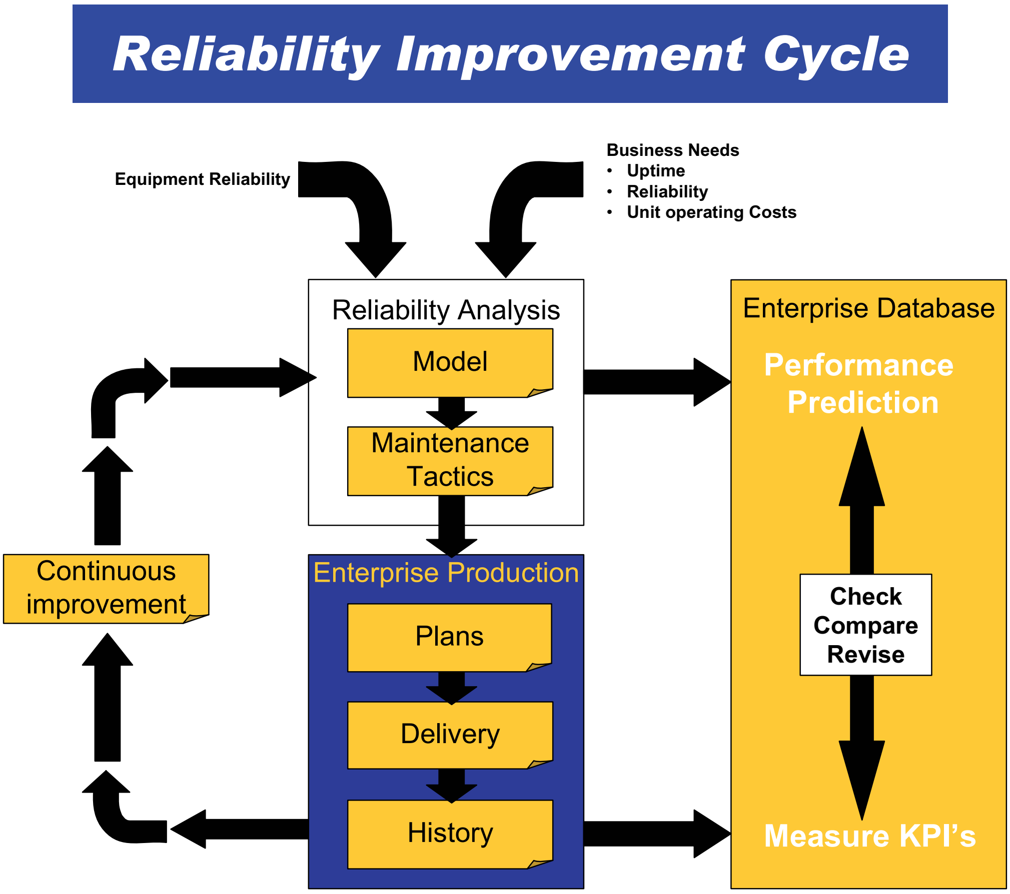


Figure 2 – Tracking Implementation Success & Continuous Improvement

Continuous Improvement

In this scenario existing EAM data is used as a basis or review. It is possible that the hierarchy structure is reviewed and updated along with maintenance strategies and task lists.

It is also likely that work order history will be loaded and analysed to help quantify maintenance strategy decisions as part of the review exercise.

Similar to above the asset hierarchy may be re-loaded along with revised maintenance strategies and task lists. In this case data will be modified rather than created.

In order for an Enterprise approach to be successful it needs to address the three scenarios above which represent the most common application of reliability studies.

4 CASE STUDY

The case study outlines below corresponds to the addition of assets – unique assets as outlined above. The reliability project provided the basis of the asset hierarchy and the maintenance strategies and task lists that were entered into SAP.

An RCM study that generated maintenance plans for a plant expansion consisting of approximately 200 items using RCMCost, was conducted over a 6 week period. 518 maintenance tasks were developed and optimized within the RCMCost program.

Typically these tasks would be grouped, configured and exported for upload using the SAP Legacy System Migration Workbench (LSMW). The tasks would also be configured into a standard work instruction document format ready for completion with detailed work instruction information. These documents would be completed manually, formatted and uploaded.

Several other tables and data would also need to be created and uploaded to SAP by an experienced SAP user

- Maintenance Strategy
- Maintenance Items
- Maintenance Plan Headers
- Task Lists
- Task List Operations
- Attached Documents

On this site several RCM studies have been completed and the implementation of the completed plans was a time consuming exercise which had actual seen the tables generated more than once to changing business rules.

The ERM generated the plans and tables in a fraction of the time estimated to configure and load the plans manually. This efficiency allows the users to focus on work instruction content rather than word document formatting etc. ERM applies consistent business rules to the generation of the strategies and Task lists.

In this particular case the site modified its business rules on the basis of other SAP improvement projects. The use of the ERM made this alteration quick and easy to reflect in the new plans, by simply updating the business rules and regenerating all the tables and documents.

5 SUMMARY

Whilst the benefits of RCM are understood by the reliability practitioners, and organisational view is naturally on results. The results of RCM studies rely on the implementation of the generated plans, then the ability to monitor and quantify performance improvements. Taking an enterprise view of reliability decisions and management ensures a high quality and efficient implementation process, whilst also providing a means to track the performance improvements against predictions.

DEVELOPMENT OF A NOVEL WIRELESS TRANSDUCER FOR MEASURING THE TRANSIENT TORQUE OF AN AUTOMOTIVE POWERTRAIN

Zhang N. [a], Chapman C. [a] and Crowther A.[a,b]

[a] *Mechatronics and Intelligent Systems Group, University of Technology, Sydney, Australia.*

[b] *Bishop Innovation Pty Ltd, North Ryde, NSW, Australia.*

Abstract: This paper presents the development of a specific wireless torque measurement system that is used to obtain the transient torque of automatic transmissions. Strain gauges are put on the surface of components that rotate and transfer power. The gauges are connected to a circular printed circuit board (PCB), which is mounted on the rotating component next to the strain gauges. The PCB contains an amplifier, low pass filter, A/D converter, microcontroller, digital RF transceiver and the power supply. The transmitted torque data is received by a stationary antenna and transceiver, which is interfaced to a PC to process, display and save the data.

The wireless torque transducer is installed at the output shaft of a 4-speed automatic transmission powertrain test facility for capturing the transient torque under various operational conditions. A number of tests are performed on the powertrain test facility with various specified disturbances such as tip-in, tip-out, 1-2 and 2-3 gear up-shifts. The transient output torque of the automatic transmission are measured and compared with those obtained from simulations performed with the estimated same operating conditions. The two sets of transient responses correlate with each other and hence demonstrate that the torque transducer meets the major design specifications.

Key Words: wireless torque transducer, powertrain dynamics, torsional vibration, gear shift quality.

1 INTRODUCTION

Fast and smooth gear shifts of transmissions used in automotive powertrains are desirable and often used as marketing tools in the car manufacturers' advertising. To evaluate the smoothness of a powertrain under various operating conditions and driver demands (eg heavy load, tip-in and kick-down gear shifts), the transient torsional response must be measured [1]. Furthermore, successful and cost effective measurement of transient torque of engine and transmission output, which is not currently available, can be used for the closed loop control of gearshifts.

Hwang et al [2] anaylsed free and forced vibration of a powertrain of an automatic rear wheel drive passenger vehicle and investigated torque converter clutch engagement shudder after applying tip-in engine torque input. Experimental measurements of the natural frequencies obtained from a powertrain correlated with those calculated. Jo et al [3] developed a powertrain dynamic model for an automatic rear wheel drive passenger car that included the dynamics of the unlocked torque converter with clutch stick-slip conditions. Again, analytical and experimental results for 1-2, 2-3 and 3-4 up shifts were presented and they correlated with each other reasonably well. The effects of activating element overlap, clutch control pressure and clutch friction were also investigated.

During acceleration and braking of a vehicle the speed ratios between rotating parts along with the powertrain change with respect to time and so the equivalent system parameters. In addition, there are multiple nonlinearities, which are often dynamically coupled with each other in a powertrain. To deal with these complexities, Zhang, el al. [4-5] developed the torsional finite elements and a procedure for formulating the integrated powertrain model, and the semi-analytical and numerical solutions schemes for dynamics of generic torsional systems and a complete powertrain with coupled nonlinearities. A novel powertrain test facility was developed for the validation of the obtained numerical simulations through experiments, both qualitatively and quantitatively [6-7]. All the major components of a vehicle powertrain system are integrated in the test rig, including a flywheel subsystem that simulates the inertia of the vehicle and a hydrodynamic dynamometer for simulating air drag. A specific wireless torque measurement system, capable of measuring transient torque while shaft rotating, was fitted within the facility.

This paper presents the development of the specific wireless torque measurement system that is used to obtain the transient transmission output torque. Strain gauges are put on the surface of components that rotate and transfer power. The details of the major components of the sensor system are provided. The wireless torque transducers are installed at the output shaft of the automatic transmission fitted in the powertrain to capture the transient torque under various operations conditions.

A number of tests are performed on the powertrain test facility with various specified disturbances such as tip-in, tip-out, 1-2 and 2-3 gear up-shifts. The transient response of the powertrain are measured and compared with those obtained from simulations performed with the estimated same operating conditions. Conclusions are drawn from the comparisons between the measured transient torque and those obtained from simulations.

2 TORQUE TELEMETRY SYSTEM

The torque measurement system used in the wireless torque sensor is described in Figure 1 as follows:

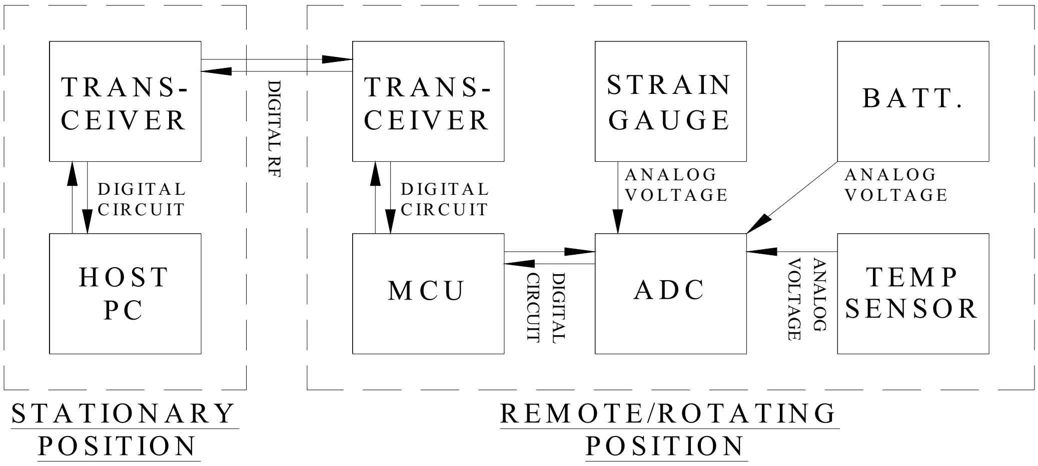


Fig. 1 Torque measurment schematic for Powertrain System

The micro-controller unit (MCU) is the Atmel ATmega8535. Its programming has been designed to take readings through its internal analogue to digital converter (ADC) upon command from the main computer. The commands are transmitted via the TR3100 transceiver. The host PC is used for controlling the remote circuit board functions. It contains an analogue section, a micro-controller, a radio transceiver, batteries and voltage regulators. The PCB take commands from the PC.

The features of the major sections of the PCB of the torque measurement system are provided below:

- Analogue section includes: a) instrumentation amplifier and filter for torque channel, with on-board shunt calibration; b) PCB temperature measurement via temperature sensing IC; c) battery voltage sensing.
- The selected micro-controller is Atmel Atmega8535, that features with a) MCU system clock speed of 7.3728 MHz; USART baud rate of 115.2 kbps; and c) ADC sample rate of 4.43 kHz.
- The radio transceiver is RFM TR3100, that features with transmitting protocol of Amplitude Shift Keyed (ASK) and transmitting frequency of 433.92 MHz.
- Three 3 Volts batteries were used and voltage regulators were used for each of the above major components.

The desired ADC sample rate is approximately 6 kHz. However due to the bandwidth limit of the RF link, the sample rate has been lowered to 4.43 kHz and every pair of samples is averaged to give a transmitted sample rate of 2.21 kHz. It is hoped that future development of the system will increase the actual sample rate to approximately 6 kHz, and use averaging to give us the maximum transmitted sample rate that the RF link will allow.

The user sends a command packet from the host computer to the remote module. The packet is in binary code sent via the radio telemetry to the micro-controller. This packet contains the user's command menu choice and the three menu parameters, one for ADC control and two for calibration timing. The packet contains a start vector at the beginning, a data separator between the vector and the data, a data separator between the data and the error checking and an end vector. The packet encoded as 12-bit binary symbols which are DC balanced in order to take full advantage of the available bandwidth. The host computer's received packet contains a start vector, 20 encoded ADC results (16 decoded results) and end vector.

3 MICRO-CONTROLLER UNIT HARDWARE

The software was written in the C programming language using CodeVision AVR. The CodeVision AVR program and compiler allows the software to be loaded directly onto the micro-controller unit through a serial port connection.

3.1 Command Functions

There are 8 possible functions within the command structure. The first 5 have been used and the last 3 are spare. Each command involves sending and receiving information from the PCB and error checking. The used five functions are briefly described below.

1. Wake command: The MCU is woken by the USART Receive Complete interrupt. Interrupts are special functions in programming that execute when necessary, rather than when called. This function also resets and starts the ADC. The wake command must be the first command function the MCU receives once turned on. It wakes the voltage regulator and sets the ADC to the 4.43 kHz sample rate.
2. Torque command: The torque readings are taken from the ADC multiplexer (ADMUX) channel 0. This channel has been set as the default channel in the program. The readings are put into a circular buffer with a maximum size of 16 samples. Then they are encoded from 10-bits to 8-bit bytes. After encoding the buffer is transmitted back to the host computer. Parameter A controls the sample rate for the ADC.
3. Temperature and voltage command: This command measures the temperature (channel 1) and the battery voltage (channel 2) on the remote PCB. A 12 microsecond delay has been included to allow for sufficient time for the ADMUX channel to change to channel 1 for temperature and then channel 2 for battery voltage. The ADMUX then returns to channel 0.
4. Shunt calibration command: For this command, readings are taken when pin B1 is set both on low and high. The readings are taken, sent back to the host PC and pin B1 is set to either high or low depending on which part of the function is being executed.
5. Reset and sleep command: This command resets the microcontroller, and puts the torque sensor into a low current sleep mode to preserve battery power. In this mode, the complete analogue section is turned off by disabling the analogue section voltage regulator, the microcrontroller ADC is turned off and the microcontroller is put into one of its sleep modes. In this mode, the only possible action that can be done is to receive and execute a wake command.

3.2 Encoding / Decoding On MCU

Encoding and decoding is needed for the ADC data and for the packetised data transmitted in both directions. The ADC data is encoded from 10-bit numbers to 8-bit bytes in order for the data to be packetised. The packetised data must then be encoded to 12-bit symbols in order to take full advantage of the available bandwidth of the transmitter.

3.3 Control of MCU Features

Universal Synchronous and Asynchronous Receiver Transmitter (USART): The USART controls the data being sent and received by the MCU. It has been set to use interrupts to process all data received and that to be transmitted. The data transfer rate between the MCU and the transceiver has been set to 115.2 kbps, as this is the fastest rate allowed by the TR3100.

Analogue to Digital Converter (ADC): The ADC has been set to free running mode at a clock speed of 115.2 kHz. Since one AD conversion takes 13 ADC clock cycles, this gives an effective sample rate of 4.431 kHz. The clock speed must be between 50 kHz and 200 kHz for maximum 10-bit resolution. The ADC is switched off when entering sleep mode so that a conversion complete interrupt does not cause the MCU to wake.

I/O Ports: Port A has been configured to be the ADMUX input (the alternate function for this port). Port B controls the voltage regulator on pin 0 and the shunt calibration for torque on pin 1. Setting the pin0 high or low turns the regulator on and off. Setting pin1 high or low toggles the step input for the shunt calibration. Port C controls the modes of the TR3100 transceiver by setting pins 0 and 1 either high or low. The setting of these pins toggles the 4 possible modes for the transceiver. Port D uses the alternate functions of pins 0 and 1 for the USART RX and TX data lines respectively.

3.4 Calibration of Torque Channel

After setting the gain of the instrumentation amplifier to suit the required torque measurement range, an end-to-end calibration was done to calibrate the complete torque measurement system. This was done with the powertrain stationary and applying a known static torque to the strain gauged powertrain component. The resulting change in torque reading is then noted at the PC.

3.5 Affect of Rotating Speed

The rotating PCB has been spin tested by mounting it on the shaft of an electric motor and driving the motor up to 6000 rpm. When this test was first performed it was noticed that the batteries moved inside their battery holders and contact was lost. An additional battery clamping clip on each battery was designed and installed. The test was repeated and no mechanical or electrical problems were observed.

4 A POWERTRAIN TEST FACILITY WITH THE STATE OF THE ART INSTRUMENTATION

A test facility with advanced instrumentation has been developed for experimental investigation of the dynamics of vehicle powertrains fitted with various types of transmissions [6]. Figure 2 shows the picture taken from the test facility and its major components include BTR 4-speed AT; Ford 6-cylinder engine, wheel hubs, tires and propeller shaft; and Spicer Axle's final drive. The facility can be used for: experimental validation of theoretical predictions of the transient characteristics of automatic transmissions during gear changes; implementation of closed loop control of gear shifts; and in particular, experimental investigations of torsional vibration of a complete powertrain caused by various nonlinearities and drive conditions.

Fig. 2 The powertrain test facility at Mechanical Engineering Laboratory of UTS.

All the major components of a vehicle powertrain system are integrated in the test rig, including a flywheel subsystem that simulates the inertia of the vehicle and a hydrodynamic dynamometer for simulating air drag. A specific wireless torque measurement system is currently under final testing. The developed wireless torque transducers are mounted at four points of interest on shafts throughout the rig. These transducers enable us to capture the transient torque at various shafts along the powertrain, such as the engine output shaft, transmission shaft, propeller and drive half shafts.

The torsional vibration characteristics of the test rig were determined and compared to those obtained from an existing car. The results showed that this test rig is a very good approximation to a real car and is thus suitable for experimental investigations of the transient characteristics of automatic transmissions. Reference [6] provides the design details of the facility and the above mentioned results.

5 TRANSIENT OUPUT TORQUE OF AUTOMATIC TRANSMISSION

The purpose of the experiment is to measure the output torque of a four speed automatic transmission during 1-2 and 2-3 gear shifts, and under tip-in and tip-out operation conditions. Comparisons are also made to those captured via the simulations. On the test rig, the automatic transmission is set in drive mode after the engine is started. The throttle position was increased and decreased gradually in order to trigger up-wards and down-wards gearshifts. The developed wireless torque transducer measures the transient output torque of the transmission during these shifts.

(1) Tip-in operation: a large throttle input was applied to the powertrain and the output torque of the transmission was taken while no gear shift took place during the short period of time. For this test, a light load was applied to the powertrain. Figure 3 shows the obtained output torque. It is noted that some data is missing in the plot, e.g., right after 2.5 seconds. This is due to the software trigger problem that exists in the microcontroller. We are currently working on an improved version of the rotating PCB that will remove this problem.

From the measured output torque, we observed that there was a significant torque oscillation after the sharp rise at about 1 second. The frequency of the torque oscillation is estimated between 6.5-8 Hz that is the same as the second natural frequency (6.9 Hz) obtained from the free torsional vibration analysis of the powertrain facility. The obtained second vibration mode indicates that the engine and automatic transmission have dominant vibration at this frequency when it is excited. Due the existence of certain damping and drag in the torque converter, the vibration was reduced significantly in the first 0.5 second.

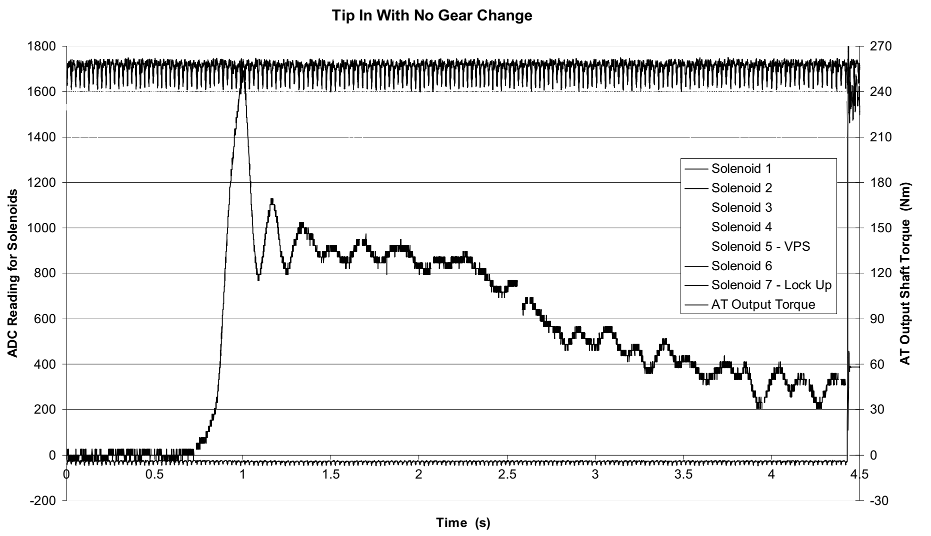


Fig. 3 Measured output torque of the automatic transmission under tip-in operational condition.

(2) Tip-out operation: after the powertrain reached a steady-state running condition with certain load, a minimum throttle input (idle speed) was applied to the powertrain and the output torque of the transmission was taken during the short period of time. For this test, a medium load was applied to the powertrain. It is noted that the torque was smoothly reduced to the minimum with small oscillations. The obtained output torque was not shown here due to lack of space.

(3) 1-2 and 2-3 gear shifts: For this test, light load was applied to the powertrain. A constant throttle position (20%) was applied and the vehicle (flywheels) was accelerated gradually and 1-2 gear and 2-3 gear up shift took place respectively. The transient output toque was captured during this period and was shown in Figure 4. From the transient output toque occurred during 1-2 gear shift, there was a significant torque hole in the torque phase and a rapid increase in the inertia phase. Similar phenomena was observed from the transient torque occurred during the 2-3 gear shift. The transient output torque during these two gear shifts was also captured from simulations on gearshift using the multi-rigid-body dynamic model of the powertrain, as shown in Figure 5 (solid line of the left plot and dotted line of the right plot). The transient output torque during 2-3 gear shift was also completed using a multi-degrees of freedom flexible powertrain model, as shown in Figure 5 (solid line of the right plot), in which significant torque oscillation occurred after the gear shift. This is partly due to a light damping in the system was assumed and the fluid drag in the transmission was ignored in the simulation. The torque oscillation is also evident in the obtained transient output torque during 2-3 gear shift shown in Figure 4 but it was reduced quickly due to significant damping in the powertrain system.

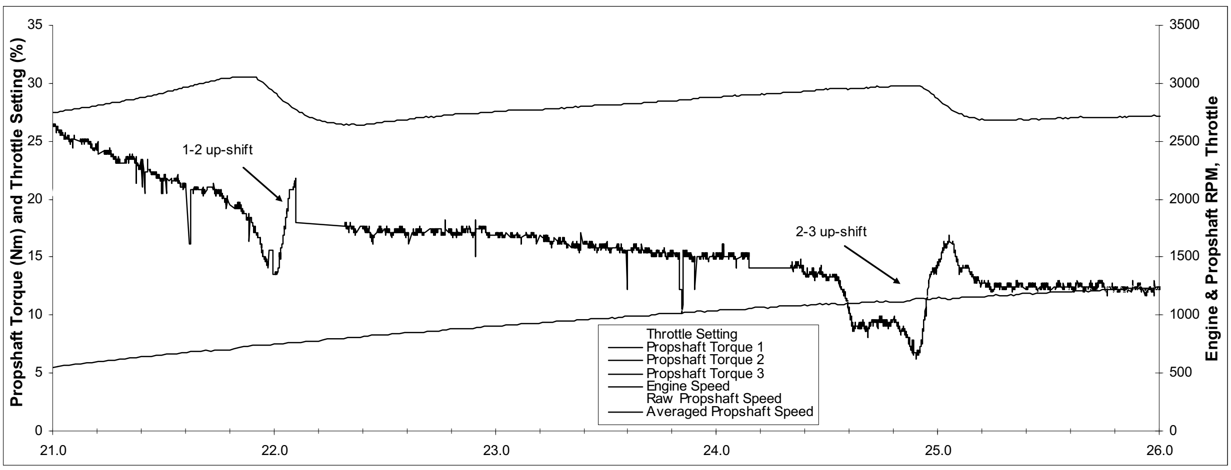


Fig. 4 Measured transient output torque of the automatic transmission during 1-2 and 2-3 gear shifts.

It should be pointed out that the torque sensor is not perfect. The data flow is periodically interrupted by a software handshaking issue, and occasionally a data packet will be corrupted by RF noise. Nevertheless, the key transient output torque during gear shifts was well captured in accordance with other measurements. Figure 6 shows the measured output torque, engine speed and averaged propeller shaft speed of the powertrain during 50 seconds of the operation.

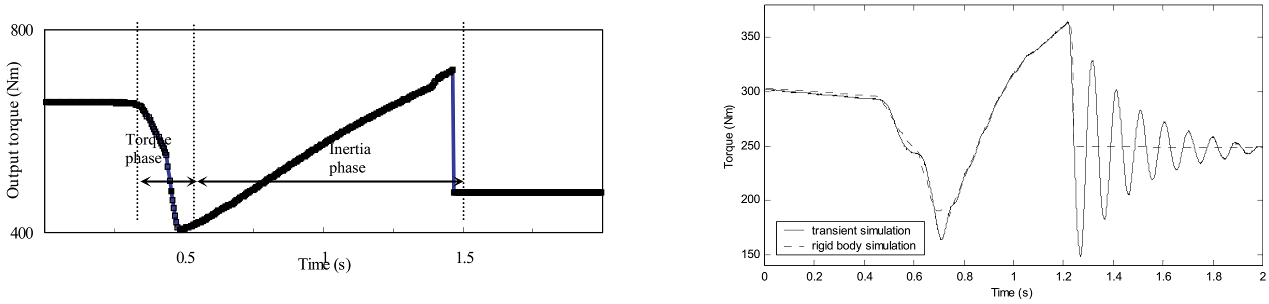


Fig. 5 Simulated transient output torque of the automatic transmission during 1-2 (left) and 2-3 (right) gear shifts.

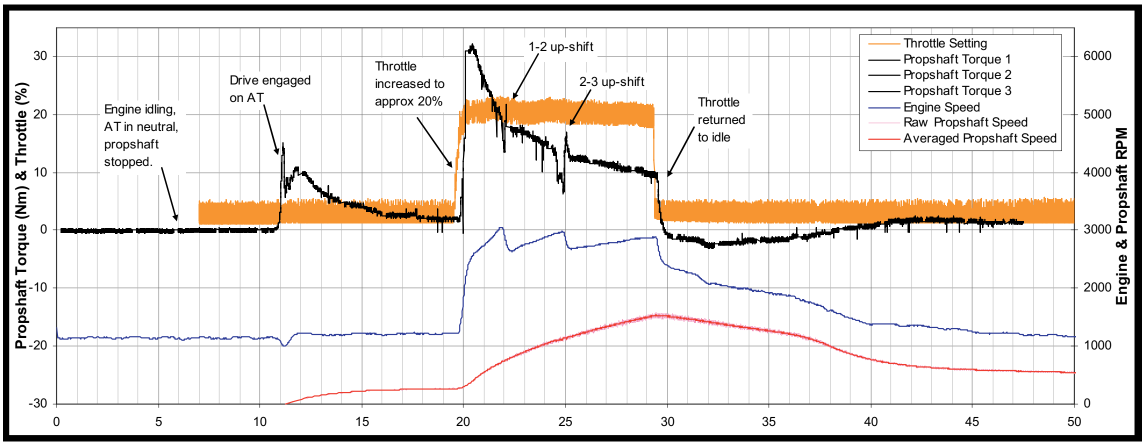


Fig. 6 Output torque of the automatic transmission and engine speed during 1-2 and 2-3 gear shifts.

6 CONCLUSIONS

This paper presented the development of a specific wireless torque measurement system that is used to obtain the transient transmission output torque. Details of the major components of the sensor, calibration and test tuning were provided. The senor was installed at the output shaft of a 4-speed transmission powertrain test facility for capturing transient torque. A number of tests have been performed on the powertrain test facility with various specified disturbances such as tip-in, tip-out, 1-2 and 2-3 up-dear shifts. The transient output torque of the automatic transmission are measured and compared with those obtained from simulations performed with the estimated same operating conditions. The two sets of transient torque correlate with each other. The experimental results show that wireless torque transducer works and meets the major design specifications. Some deficiencies existing in the current senor were also discussed.

7 REFERENCES

[1] N. Zhang (2005) Dynamics of a powertrain with ratio-dependent system parameters and multiple nonlinearities, *Proceedings of the 11th Asia-Pacific Vibration Conference,* 23-25 November, Langkawi, Malaysia, pp. 35-48.

[2] S.-J. Hwang, J.-S. Chen, L Liu and C.-C.Ling (2000) Modelling and Simulation of a Powertrain-Vehicle System with Automatic Transmission, *Intl. J of Vehicle Design*, Vol.23, No.1/2, 145-160.

[3] H.-S. Jo, et al (2000) A Study on the Improvement of the Shift Characteristics for the Passenger Car Automatic Transmission, *Intl. J of Vehicle Design*, Vol.23, No.3/4, 307-328

[4] N. Zhang, A. Cowther, D.K. Liu and J.M. Jeyakumaran (2003) A finite method for the dynamic analysis of automatic transmission gear shifting with a four-degree-of freedom planetary gearset element, *Proc. of IMechE, Vol. 217, Part D: Journal of Automobile Engineering*, 461-473.

[5] A. Crowther and N. Zhang (2005) Torsional finite elements and non-linear numerical modelling in vehicle powertrain dynamics, *Journal of Sound and Vibration,* Vol. 284 (5), 825-849.

[6] M. Wang, N. Zhang, C. Chapman and J.M. Jeyakumaran (2002) Design and torsional vibration analysis of a complex vehicle powerrtain system test rig, *Proc. of 5th Intl. Conf. on Vib. Eng.*, Nanjing, China, 303-308.

[7] A. Cowther, R. Singh, N. Zhang and C. Chapman (2006) Impulsive Response of an Automatic Transmission System with Multiple Clearances: Formulation, Simulation and Experiment, *Journal of Sound and Vibration* (in review).

CUT YOUR COSTS BY 75% - HOW A LITTLE PREVENTION CAN SAVE A LOT OF MONEY

Allan Bruce

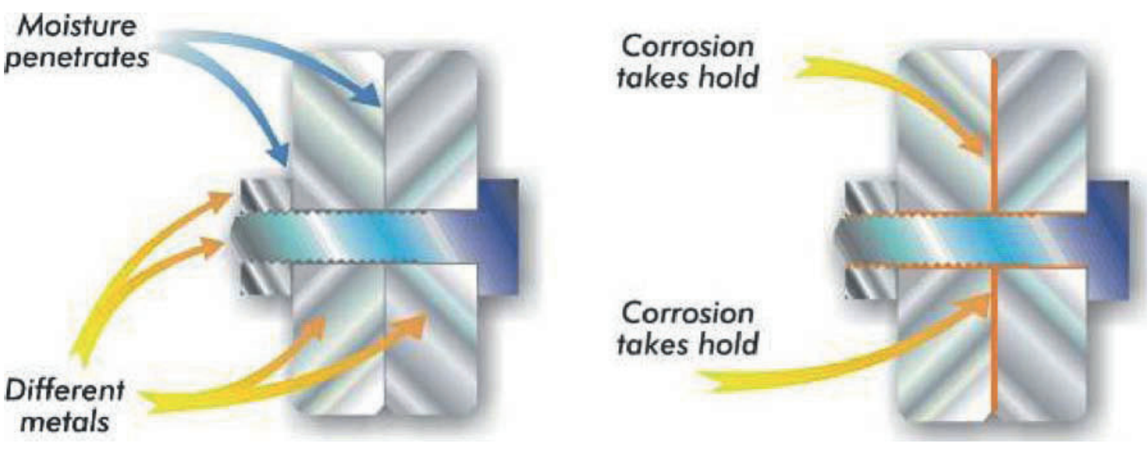
The Alocit Group, 112 Stirling Highway, NORTH FREMANTLE WA 615, Australia

Abstract: An introduction to cost benefits of the Enviropeel System, a system that uses a heating and pumping system to melt and spray a thermoplastic material in order to provide a corrosion-inhibiting protective coating around substrates of any size or shape. The material contains its own inhibitors, thus providing both passive and active protection for protected surfaces. As part of the development process, long-term testing of Enviropeel forms part of the DNV JIP Bolt Testing Programme, a programme which includes Conoco-Phillips and BP, at whose behest the tests were planned. First results after 18 months of the 10-year programme show very good results for Enviropeel. Case studies have shown from various areas around the world with a focus on the experience of the Enviropeel System in WA with BHP Billiton and Dampier Salt. Results over the past three to four years have been so spectacular, with projected lifecycles of equipment increased by a factor of 5, that the Enviropeel system has become standard practice due to the substantial cost savings.

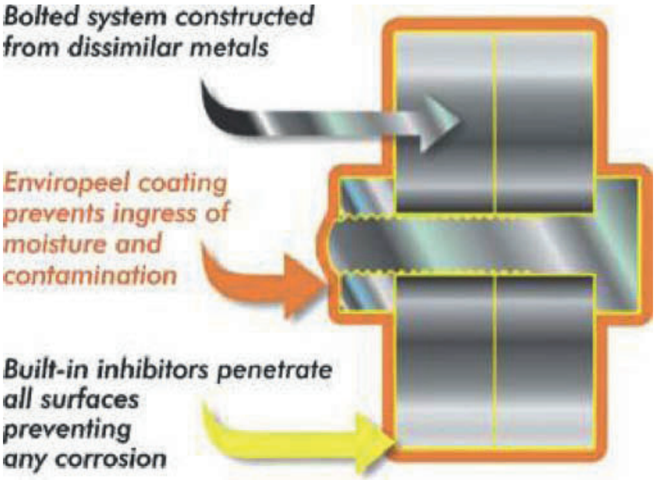
Key Words: Alocit, bolts, bolted systems, cost effective, component life, corrosion-inhibiting, dipping, , encapsulating, environment, Enviropeel, loading, mining, offshore, polymer, pulley, recyclable, renewability, re-usable, Robil, thermoplastic, samples, Slugger, flanges, valves, sprayable, Pilbara, conveyor bearings, BHPBilliton, Dampier Salt

1 INTRODUCTION

The Enviropeel system uses a specially designed unit to spray apply a corrosion-inhibiting thermoplastic barrier coating. It was developed to provide a solution for the many corrosion problems associated with bolted systems such as flanges and valves, by adopting a whole system approach rather than providing a variety of individual solutions for the wide range of problems that beset such systems.



When flanges or valves fail both safety and cost implications can be quite severe. At the most mundane level, replacement will probably involve costly shut downs, cutting bolts and refurbishment. Leaks can affect the environment and the potential for severe consequences is ever present. Bolted systems are vulnerable to corrosion in a number of ways. From galvanic corrosion because they contain a variety of components, often of quite different materials – stainless steel flanges with carbon steel bolts, for example. Gaps and voids bring pitting corrosion, assembly damages coatings and washers often fail.



Enviropeel isolates and encapsulates the substrate, depriving corrosion of the moisture it needs and bathing the entire system in corrosion inhibitor, actively preventing rust and ingress of contamination.

2 THE STATUS QUO

What levels of protection are provided by current systems? In truth, not a lot. Of course, you can keep cleaning and painting on a very regular basis but, at best, this locks all the bolt with accumulated paint and often leaves flanges streaming with rust in as little as 6 months. Flange protectors, tape wraps, bolt caps – all can play a part but are either partial solutions or so lock away the protected substrate that access is extremely difficult when required.

Pictured left is an example of what can happen despite the best efforts of engineers to provide protection.

Although caps have been used to protect the bolts - they are not protecting anywhere else - and they are not even possible to use on all the bolts because in some areas the heads are so close together access is impossible. Caulking has been used to try and protect the flange faces – but it has shrunk back, creating the perfect environment for moisture to get in and corrosion to accelerate!

And you have to feel sorry for the allen bolts at the base – they have no chance. Streaming rust from the structure above, galvanic corrosion from the various metals in the system – and a perfect collection point in the top of each bolt, apparently designed to collect water and generate corrosion!

3 WHAT DOES THIS HAVE TO DO WITH AUSTRALIA?

The flanges above show what we believe to be the best available protection for bolted systems – completely protected from its environment and actively protected by built-in corrosion inhibitors.

The pictures above show typical examples of what they were trying to prevent. Believe it or not, the picture on the right is a brand new pulley – or at least it was before two years of storage – the wrapping has failed and there is significant corrosion damage in the bearing

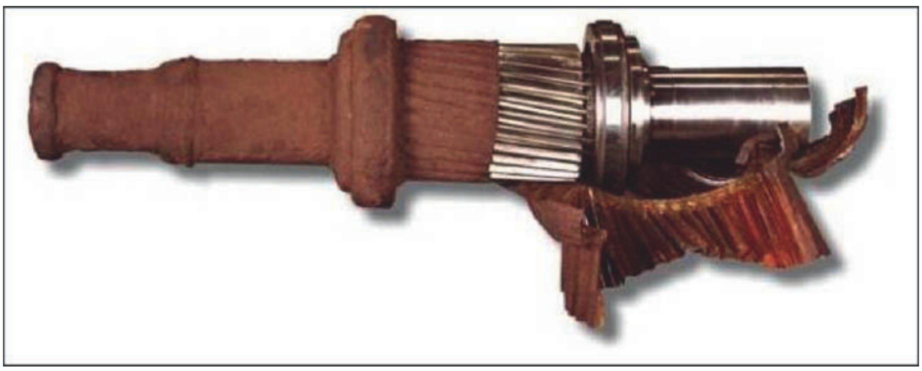
Enviropeel was first introduced into Australia by a director of an engineering company based in the Pilbara. The company, Robil Engineering, was heavily involved in work for the mining industry in the region. As the work involved extensive re-engineering, with lengthy storage periods before installation, corrosion, both before and after the engineering work, was a major cost factor.

The ability to provide a dual purpose coating – providing both active and passive protection – to any object, no matter its size or complexity, was what attracted Robil to Enviropeel.

Robil began by applying Enviropeel to machined surfaces and components for storage. Soon, they were applying it to complete assemblies - and finding that all their previous problems with corrosion during storage had been eliminated.

Soon Enviropeel was being applied to all machined parts and steel assemblies – and achieving remarkable results. The gear shaft below typifies the results that Robil found –peeling away its protective coating after two years exposure in aggressive conditions, the gear shaft emerges in fantastic condition.

With such success with stored parts, the Company's thoughts turned to what could be achieved on operational equipment. One of the major components that Robil had to re-engineer was conveyor bearings. Hundreds of bearings from the many miles of conveyors operated by major customers BHPBilliton and Dampier Salt required constant replacement because of the high levels of iron-ore dust and salt in which they had to work.

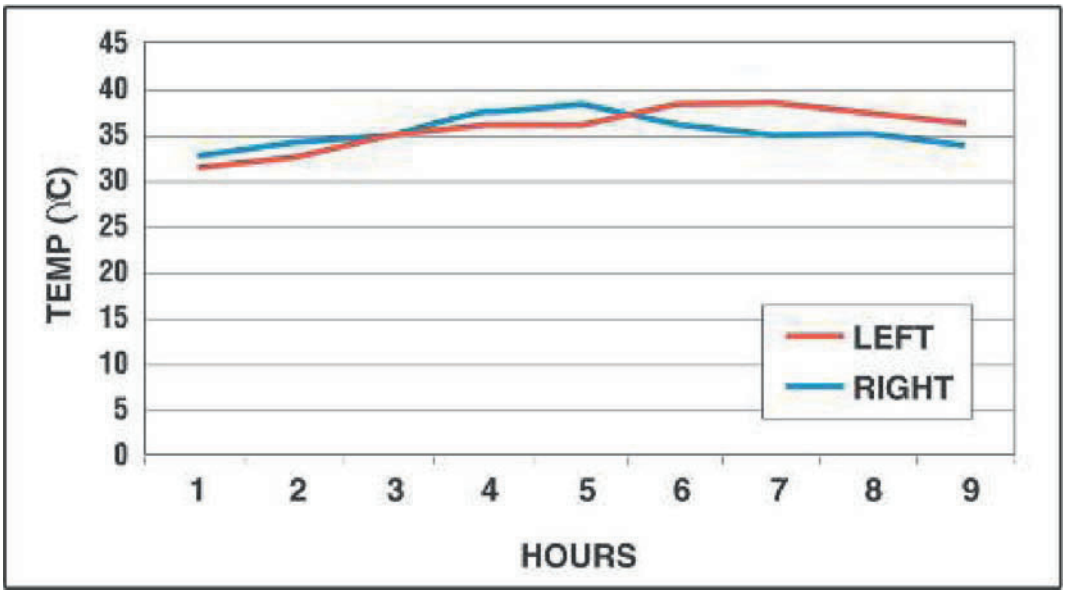


This bearing race on the right shows the wear patterns that produced failure

It's easy to see why Robil thought something needed to be done. The condition of these components is typical of what could be found in the most aggressively corrosive areas

Because of Enviropeel's ability to be sprayed even on to the rotating shaft of the pulley, it seemed to Robil that, not only would the material provide corrosion protection but that it might also prevent ingress of the dust and other contaminants that were so damaging to the bearings.

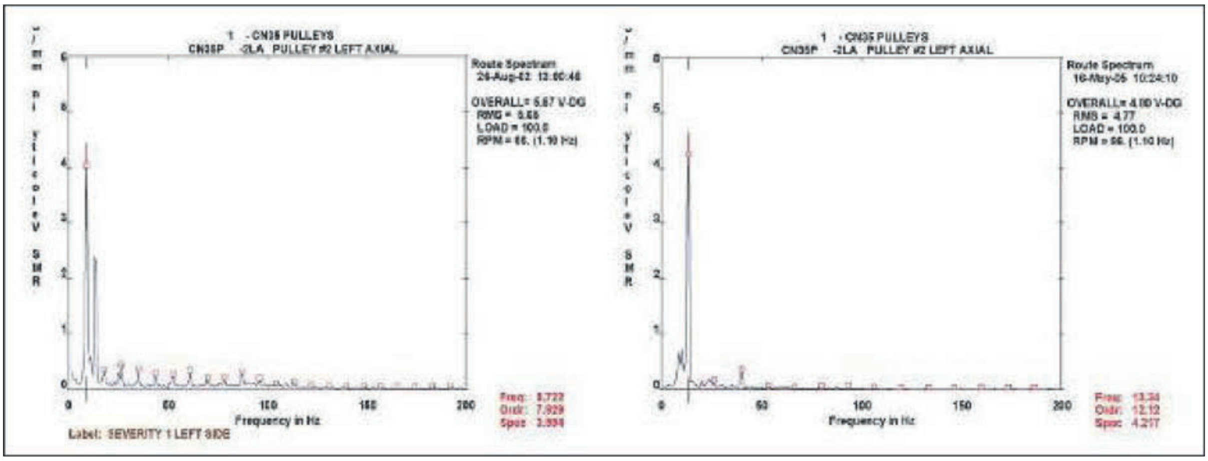
In the most severe areas, pulley bearings, which were costing $8000 to replace, were failing in an average of only nine months. If something could be done to increase their lifespan, the savings to the mining companies could be considerable.



Tests first had to be undertaken to see if the coating would have any adverse effect on bearings while in use – would the material increase operating temperature or prevent access for maintenance or inspection. Tests showed that the coating did not increase running temperature and access, where necessary, was easily achieved and the coating re-applied.

With cooperation from BHP Billiton and Dampier Salt, Robil installed test bearings with their housings still encapsulated in Enviropeel and the trial process began. It soon became evident that the system was a winner. Following the introduction of Enviropeel, warning indications from the bearing monitors ceased.

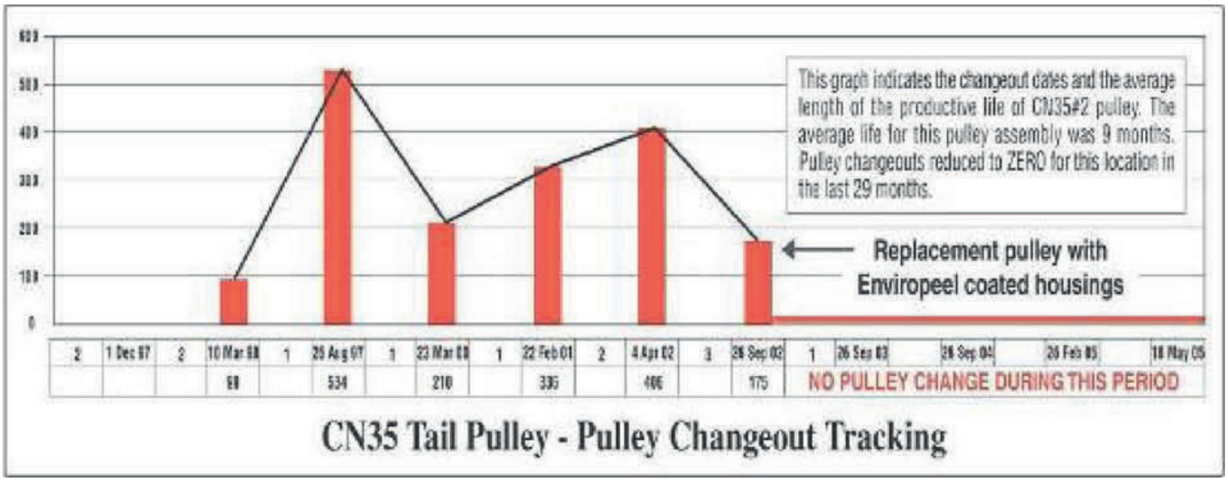
Conveyor operators monitor the state of the bearings for vibration and spectrum graphs, like these shown below, measure the vibrations recorded as wear occurs and bearing clearances increase. On the left-hand graph unprotected pulleys show typical large measurements as wear increases to the point of failure, whereas on the right the consistent low signals following the introduction in Sept 02 of a pulley coated with Enviropeel, during which time no further replacement of bearings has been required.



Note: The spike seen on the left of this particular graph is the lower frequency screen signal picked up during readings taken that particular day, and does not relate to wear factors. The earlier graph, on the left, shows a vibration level with a Severity 1 rating, indicating urgent change-out is needed to avoid a potentially dangerous failure.

Severity ratings commence at 4 (lowest) and progress through to 1. The more aggressive the environment, the quicker the wear rate and severity upgrades on unprotected pulleys.

This data shows what a difference Enviropeel makes – prior to September 2002 bearings were failing at an average rate of 9 months in the severest areas. Once Enviropeel was introduced, no further replacements were required – no excessive vibrations, no severity warnings. The graph shows up to May 2005 but in February 2006 there have still been no failures



The contrast between the corrosion on the roller in the top left of the picture and that of the bearing housing – both of which were refurbished at the same time - shows how effective Enviropeel can be.

4 COST SAVINGS

The potential for cost-savings is enormous. The system is currently showing an increased lifespan of over 400%. On one pulley alone, over 5 years, the saving will be over $50,000 and, although not all pulleys are in such severe areas, there are over 4500 pulleys in the Pilbara alone, with an estimate of at least 15,000 across Australia. Whatever the potential lifespan of the various bearings, it is obvious that millions of dollars could easily be saved if they were all using Enviropeel.

5 CONCLUSIVE RESULTS

An outline of the results from Robil makes it clear why their innovative approach won them a 2005 Engineering Excellence award.

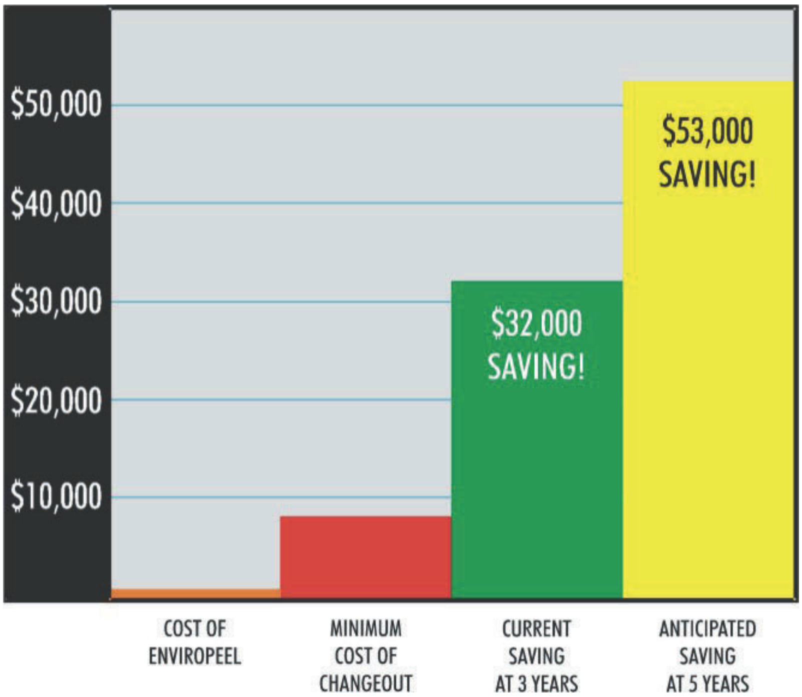
5.1 On stored conveyor pulleys

- Return for replacement without Enviropeel 44.5%
- Failure rate with full Enviropeel protection 0%

5.2 On operational pulleys

No bearing has failed since Enviropeel applied, previous average only 9 months!

Cost saving per bearing



Average bearing life in original trial location	**9 months**
Current bearing life in original trial location with Enviropeel applied	**36+ months**
Resulting component life increase	**400+%**
Resulting saving in pulley change out costs	**$32,000+**
Reduction in maintenance costs	**95 %**
Percentage of Enviropeel costs to rebuild costs	**10-15 %**
Percentage of Enviropeel costs to pulley change out costs	**5-7 %**
Resulting percentage reduction in risk exposure	**63+ %**
Anticipated increase in component lifetime	**670 %**
Average reduction in greasing regime	**75%**

6 CONCLUSION

Enviropeel is environmentally friendly. Reusable and recyclable it saves waste and contains no VOCs or other toxic materials.

It is easy to remove, when access is required and can as easily be reapplied – even using the same material!

We have already seen what cost savings can be made but the system, by reducing the amount of maintenance and risk involved in such complex renewals, also brings significant safety benefits.

Already specified by BHPBilliton and Rio Tinto for use on their equipment, Enviropeel is also being used or assessed by a number of other companies in all parts of Australia.

Enviropeel is a simple product with a huge market for overcoming ingress and corrosion of critcal spares and equipment.

The savings to date have been calculated to be in the thousands of dollars. What hasn't been mentioned is the LTI's or injuries prevented due to the reduction in the number of change outs required, which would be well over 75%!

MEASURING SAFETY IN GAS DISTRIBUTION NETWORKS

Ype C. Wijnia [a,b] and René J.M. Hermkens [c]

[a] *Researcher, Division Energy and Industry, Faculty of Technology, Policy and management, Delft University of Technology, Jaffalaan 5, Delft, The Netherlands(Y.Wijnia@tbm.tudelft.nl)*

[b] *Risk Manager, Essent Netwerk, Pettelaarpark 80., 's Hertogenbosch, The Netherlands.(ype.wijnia@essent.nl)*

[c] *, Senior Consultant, Kiwa Gastec Technology, Wilmersdorf 50 7327 AC Apeldoorn, The Netherlands (Rene.Hermkens@kiwa.nl)*

Abstract: In the Netherlands, the safety of the gas distribution system is a hot issue. The public perception is that the safety of the system is decreasing, whilst the industry thinks that it is business as usual and that neither the number of incidents nor the consequences of incidents are larger than they used to be. Which of the claims is true? To this question there is no objective answer. First of all, no accessible database exists that contains records of all relevant incidents. Secondly, there is no agreed upon method to combine all those incidents into a simple, easy to understand indicator of safety. In this paper we first explore the theoretical framework for such an indicator. Combining this framework with actual incident data will result in a quantitative model. Finally, we present the results of the quantitative model. The paper ends with an outlook for the application of the indicator in the quality regulation of the Dutch gas distribution companies.

Key Words: Safety, measurement, indicator, gas distribution,

1 INTRODUCTION

The gas distribution industry in the Netherlands claims that the system is safe. To support this claim, they refer to countless numbers of NEN/EN/ISO standards, which are applied in the design and maintenance of the system. Phrased differently, they state that gas distribution is highly technical, that experts have thought about the details thoroughly, and that those experts should be trusted. As both the standards and the experts are from within the industry, the basic message is: "You can trust us, we are the experts, we know and we care", a message that worked well for a long time.

In the current society, however, this asked for trust does not come naturally, as Shell, for example, experienced with the Brent Spar. In short, the Brent Spar was an oil storage rig in the North Sea that had to be decommissioned because its function had been replaced by pipelines. A Best Practicable Environmental Option study showed that the best way forward was a deep sea disposal. Greenpeace, however, did not agree with this study. They stated the volumes of toxic materials were much higher than Shell claimed and possessed the rig to prevent the disposal. The public assumed Greenpeace was more trustworthy than Shell and boycotted the Shell gas stations, an act which forced Shell to change course. In the aftermath, Greenpeace had to admit that they were wrong and Shell was right about the toxic content of the Brent Spar, but that did not help Shell anymore. The option of deep sea disposal was definitively out of question.

In the debate on the safety of the gas distribution system, some parallels can be drawn to this Brent Spar example. First of all, the liberalization of the Dutch energy markets in the period 1999-2004 resulted in lots of faulty invoices from the energy companies, destroying the image of a trustworthy industry. The large 3 energy companies ranked high in customer dissatisfaction (top3), according to consumer watchdog television programmes like Kassa and Radar. Secondly, because of this bad press, every gas related incident was connected to the liberalization, even if it had nothing to do with the gas distribution system. An example can be found in the explosion in The Hague (June 28, 2003), which according to the first reports was caused by gas leakage from a gas pipeline, where in fact it was caused by a leaking propane tank in the cellar. As a bonus, the public awareness of incidents was further enhanced by the rise of numerous reality TV shows and local human interest programs on the increasing number of nation wide television stations. The odds of an incident drawing attention of one of such shows were further enhanced by new technologies like the internet and cellular phones, speeding up the propagation of news. Furthermore, the Dutch safety board has published a few reports with some very alarming conclusions. As a result it seems as if the number of incidents is increasing.

In this light, it is not very surprising that the public did not take the industries claim of safety for granted and demanded additional measures. The awareness of incidents increased, and awareness drives the perception by the availability bias (Tversky and Kahnemann, 197x). But is there any truth in the perception? Is the gas grid really becoming unsafe?

Embarrassingly, to this question no objective answer existed. Beside the statistics on leakage and the standards, the industry did not have any facts supporting the claim of safety. On the other hand, there was no evidence for an increase in the number of incidents either, but the industry did have to prove its claim as perception was the public truth.

To issues revolving about the public perception, the best approach is not to claim expertise and declare safety, as the industry was used to (and Shell did initially with the Brent Spar), but to share knowledge and data (Shells reprise). This created an interesting challenge for the gas industry, as they did not have very much knowledge and data to share. First of all, they did not have an easy accessible database on all incidents and accidents related to the gas network. Secondly, they did not have any methods to translate such a list of incidents into an easy to understand measure of safety.

Fortunately, some experience on quality issues existed. Most gas distribution companies in the Netherlands also distribute electricity, and for the electricity grid, the major quality indicator is Customer Minutes Lost, which had been recorded for some 25 years in the so called "Nestor Enquete". If such an instrument could be developed for the safety of the gas grid, it might prove helpful in the public debate. However, discussions on safety issues are always difficult, as extensive research on this subject has shown (Slovic, 2002). Furthermore, even the existence of an objective metric for safety does not guarantee the right perception, as the reliability indicator for the electricity grid shows. The perception is that electricity is interrupted once every year for about 2 hours, where in fact it is only once every 4 years according to the 25 year statistics (SEO, 2004).

Despite these potential barriers for acceptance, just knowing for sure was tempting enough for the industry. Both authors were asked to take part in the development of the safety indicator. In this paper, we present the results of this development process. We first address the issues surrounding the concept of safety, like actual figures versus disaster potential, adding different types of accidents, the value of a human life and so on. This is the value judgment section. The next step is building a conceptual model for the indicator which translates the measured inputs into a single metric. As we shall see, a direct assessment of the safety consequences is not a reliable indicator for the safety of the gas distribution system. A deeper understanding of the underlying mechanism is needed. This is the conceptual part. Finally, we will fill this conceptual model with actual incident data, resulting in a quantitative estimation of the safety risk in the gas distribution system. We will end the paper with conclusions on the usefulness of the indicator.

2 THEORETICAL FRAMEWORK

In this section we address the issues surrounding the debate on safety. As we will see, much of the issues require some kind of value judgement on a certain aspect of the total safety abstraction. To structure the discussion, we will first focus on the different concepts of safety. Bases on the characteristics of the safety risk and the uncertainty involved, we will choose a concept for the indicator. Next, we will address the issue how to count safety incidents. Safety incidents can have quite different consequences, like personal injuries or fatalities, but also property damage. Adding those consequences is not trivial. This non triviality holds within a consequence level (is every fatality equal), between different consequence levels (how much worse is a fatality than a serious injury) and between different values (how should a fatality be compared with financial losses, or customer minutes lost). We will end up with a valuation scheme for the different types of consequences.

2.1 The concept of safety

Safety can be used in a myriad of ways. Sometimes it refers to financials (this is a safe investment), sometimes to information (how safe is the internet), but in many cases it refers to personal injuries (road safety figures). Furthermore, safety can be about the actual figures (the annual fatalities on the road) or about the expected figures, the so called safety risk (e.g. nuclear power). Discussions on actual figures revolve around the definitions. For example, does the road safety figure only count the victims at the crash site, or are the ones in hospital afterwards included? Until how many days after the accident should those be counted? Those issues are solved relatively easy.

Discussions on safety risks are much more difficult. This is because different views exist about the essence of risk. In the first conceptual approach, risk is seen as an entity that can be objectively measured or calculated. In this view, risks are typically expressed as a product of consequence and likelihood. This objective risk is used by financial managers, safety engineers, decision analysts and so on.

The second approach, challenges the assumption of objectivity, as it is not this objective risk that determines if a risk is acceptable. Some people smoke (high objective health risk) but worry in the same time about electromagnetic radiation from High Voltage Transmission Lines (health risk uncertain). Research by Slovic has shown that this acceptance level is driven by so called psychometric factors (figure below) and not by the objective risk level. Risk is therefore very subjective.

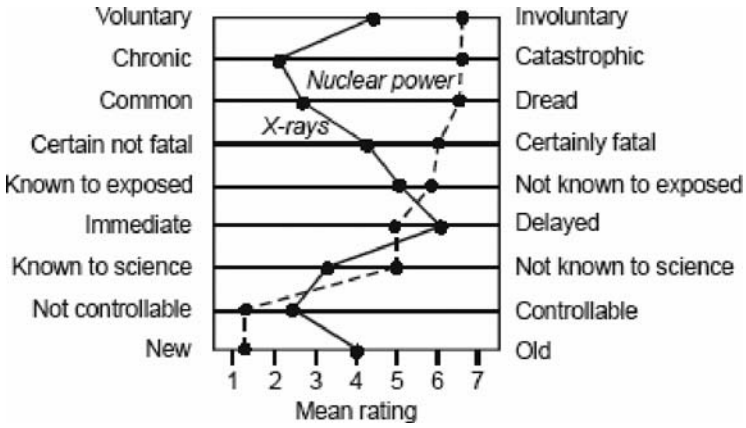


Figure 1: Psychometric factors for risk perception (source Slovic and Weber (2002)

Despite the usefulness of this subjective risk in risk communication, it does pose some problems for risk managers. For example, if the acceptance of a risk is determined solely by its disaster potential (like it was the case for nuclear power), no amount of risk measures can make it acceptable, as the disaster will always be thinkable. However, stopping all activities with a disaster potential might reduce the standard of living by a considerable amount, a sacrifice that few citizens are willing to make.

The constructive approach tries to bridge the gap between the objective and the subjective concepts, by stating that a risk is a social construction that human beings have "invented" to deal with the dangers and uncertainties of life. The dangers and uncertainties are real, as is the response to the risk (Thomas Theorem), but the risk itself does not need to be real. This approach is advocated by the USA's National Research Council (NRC, 1996). A risk can therefore be defined as *a potential problem that needs to be decided about*. This view is further supported by the value judgements needed to arrive at a risk figure (as mentioned in the introduction to this section).

Which of those concepts should be used for the safety indicator? For answering this question the classification scheme of Klinke and Renn (Klinke and Renn, 2002) could be used. They propose 4 different kinds of decision approaches for different risks, being rule based, risk based, discourse based and the precautionary principle. The drivers in this classification scheme are uncertainty, disaster potential and the social mobilization factor. As the uncertainty is pretty low (over 50 years of experience), the key issue is whether or not the gas distribution system has disaster potential. Natural gas can cause large explosions (e.g. Alpha Piper) but this is only possible at pressures well above the ones used in the gas distribution system (EGIG 2004). Therefore, a risk based approach could be used, even though some value judgement is needed. To manage the potential for social mobilization it is best to make it a collaborative effort by the industry and the regulating bodies.

2.2 Comparing incident types

In order to be able to add different types of accidents into a single number, some kind of a valuation scheme is needed. This valuation scheme should answer three questions: Which are the affected values that will be taken into account? What levels of severity will be distinguished? What are the equivalent values for the different levels of severity?

Those questions are not unique for a safety indicator. On the contrary, most companies familiar with risk management use a valuation scheme in their risk matrices. What we are looking for would be a risk matrix without a probability axis, being a consequence matrix. In case of the safety indicator, the only question left to answer would be what affected values would be taken into account. The use of personal safety would be trivial, but was any other value needed? The obvious candidate would be property damage, a well known consequence of gas explosion. Property damage could be assessed objectively. Therefore its inclusion would not threaten the objectivity of the indicator. In the table below the severity of different consequences for those values is shown.

Table 1: Severity levels of personal injuries and damages

Severity level	Personal injury	Damage
6	Multiple fatalities	> 10 million euro
5	Fatality	1-10 million euro
4	Serious injury	100k-1M euro
3	Lost time incident	10k-100k euro
2	Near miss/first aid	1-10 k euro
1	Unsafe situation	< 1k euro

In this matrix the logarithmic character of the severity classes is very clear for the value “property damage”. Quite surprisingly, this logarithmic character also holds for the personal injuries. Research has shown (iceberg theory) that for every fatality about 10000 unsafe situations, 1000 near misses, 100 lost time incidents and 10 serious injuries occur. Even though adding injuries is quite different from adding damage (financial damage can be transferred, personal injuries cannot), if used with the mentioned scaling factors it makes sense statistically. It allows us to express any combination of safety incidents in the equivalent fatalities. The scaling factor between the values seems to be an industry standard (Shell 2002). By expressing even the injuries in the monetary equivalent, all incidents can be expressed on a single scale, although this does in no way mean that a human life can be replaced by a certain amount of money. After all, no markets exist in which people sell their own lives.

3 DIRECT ASSESSMENT OF THE SAFETY INCIDENTS

Now that we have defined what we mean by safety, we can start counting. For this we used the Kiwa/Gastec accident database, which holds records of all network related accidents in The Netherlands dating back until 1993. In addition we used the reportable incidents of the Dutch safety board, although they only kept records for 2004 and 2005. The distinction between accidents and incidents is not meaningless. An accident is an event in which someone got hurt or third parties property was severely damaged (> € 500,000) as a direct result of the event, whereas an incident is an event in which this might have happened. The cost of the measures taken to prevent an incident becoming an accident (like evacuation costs) are not included. The total number of real accidents was 39 over the period 1993-2003, which averages 3-4 accidents per year. However, the variance is very large, as shown in the figure below. The number of accidents varies between 2 and 5, but the equivalent value ranges from almost zero (1996) to over 10 million (2002). Clearly, no correlation exists between the number of accidents and the equivalent value. Besides, it is highly unlikely that the outcome is correlated to the quality of the grid. This means a metric based on a direct assessment of the accidents does not produce a robust outcome if it is to be used for investment decisions.

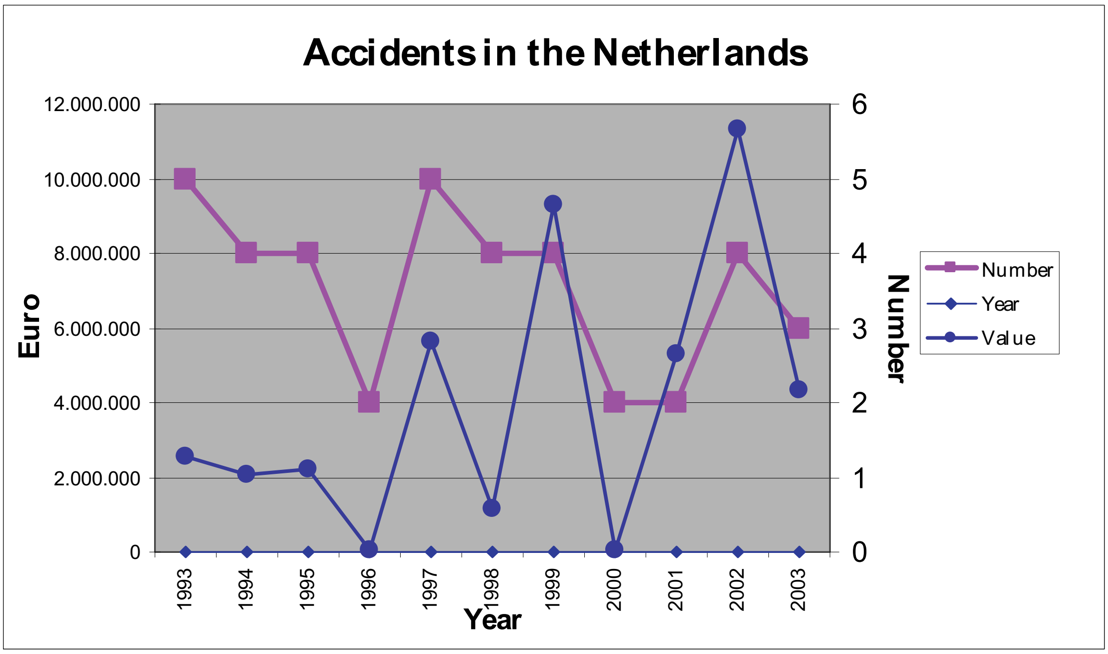


Figure 2: Gas related accidents in The Netherlands, number and equivalent value

This lack of robustness is a result of the low number of accidents, on average about 3 per year. As the indicator is meant to help in investment decisions, lack of robustness would be a fatal flaw. To overcome this problem, the data used to calculate the safety has to be extended. We will use the incident process (based on Wijnia and Herder 2004) to get a better understanding what those data should be.

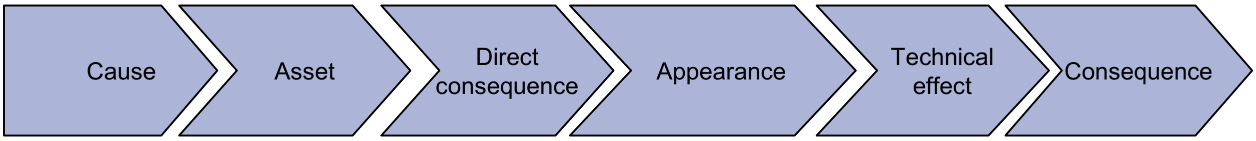


Figure 3: Incident process

The incident process states that value consequences do not appear out of nothing, but are a result of a chain of cause and effect. For example, excavation works (cause) can damage a service line (asset), resulting in leakage (direct consequence). The escaping gas might accumulate (appearance) in a closed space. If the gas explodes (technical effect), it might destroy property and injure people (value consequence). Barriers might exist between the different phases in the incident process. Not all excavation works damage pipelines, not all damaged pipelines leak and so on. This means that if we can reconstruct the processes that led to the mentioned 39 accidents, we might be able to measure safety based on situations occurring earlier in the process. As the number of those situations is much higher than that of real accidents, it is more likely to produce a statistically sound figure.

This incident process based approach also matches the concept of safety as a risk, as it addresses not only the real accidents, but also the potential accident. Applying the incident process to the recorded accidents led to the conclusion that only a few combinations of causes and assets, of the thousands possible ever led to a safety accident. The causes and assets are shown in the incident triangle below.

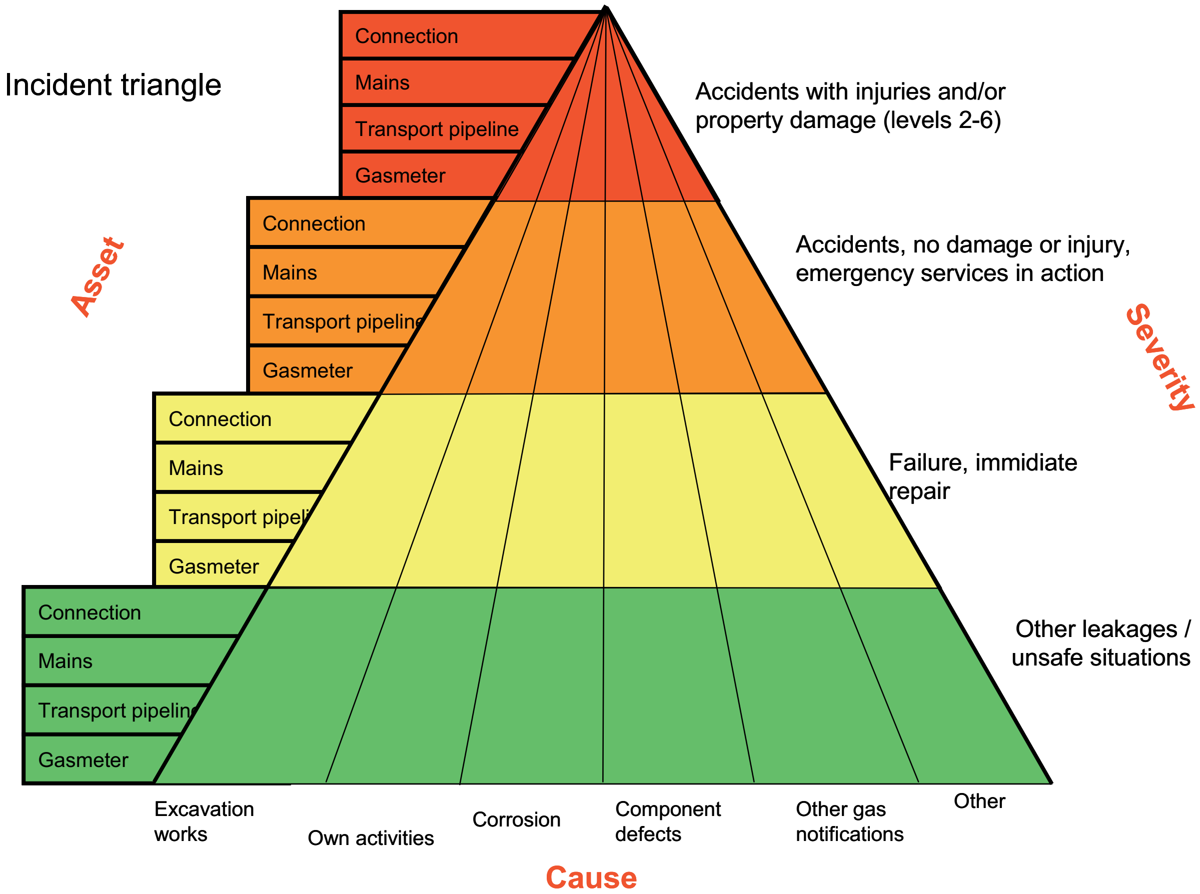


Figure 4: Incident triangle

A direct link exists between the incident process and the triangle. For each combination of asset and cause, four levels exist on which an incident can end, corresponding to the phases 3 to 6 of the incident process. The top level (red) accidents are those incidents that completed the incident process and had a value consequence. The second level (orange) represents the incidents that showed technical effects, but did not create heavy value losses. The third level incidents (yellow) are the ones that did have a public appearance, but no immediate dangers. Finally, the bottom level (green) represents those incidents in which a cause had a direct effect on an asset. Situations in which the asset was not damaged by the cause are not considered to be incidents.

In the incident process, a limited likelihood of propagation exists. Only a fraction of the assets will be exposed, only a fraction of causes acting on assets will create damage, only a fraction of the damage will reveal itself by escaping gas, only a fraction of the gas clouds will be within the explosion limits, and finally, not all dangerous situations will actually damage property or hurt people. Therefore, the number of incidents in the green level is much higher than the red level. This is in line with the iceberg model as described by Heinrich (1931). An indicator based on green level incidents should in theory be more reliable than an indicator that measures top level accidents. In practice, however, this does not hold, as the reliability of the indicator also depends on the quality of the incident registration, which is far better at the top level than at the bottom level. Fortunately, in the Netherlands we have Nestor-Gas, a database which holds records of incidents that required immediate action (yellow). This satisfies both the need for a robust number of incidents as the need for a reliable data source, and therefore will be used as the basis for the indicator.

With the framework presented above it is possible to calculate the total risk of a group of assets. However, absolute figures often do not mean much, as it does not tell about the size of the population from which the figures were derived. For example, road safety is always expressed as the number of fatalities per 1000 inhabitants, or per million miles. This makes it possible to compare different years within a country, or to compare different countries. Which metric is the correct one, however, depends on the question to be answered. The inhabitant based metric tells about the likelihood of an average person to die from a road accident. If the question is about the safety of road design, it is better to use the mileage based figure. For the safety indicator, both an inhabitant related metric and a network length related metric can be used. Because the issue is public safety, it seems logical to use the inhabitant based figure. So, to calculate the safety indicator for a group of assets, we have to count the number of different types of incidents, multiply those numbers with the average incident risk, add the risks and divide the total by the number of connections in the asset group.

This concludes the theoretical framework of the safety indicator. We have defined what we mean by safety, and specified a methodology to add all accidents into a single figure. In short, the safety indicator is a metric that is proportional to the likelihood of an individual to die from an accident in the gas distribution network. In the next section we will look at the quantitative properties of the indicator.

4 THE QUANTITATIVE MODEL

In the previous section we have explained the methodology of the safety indicator. Now it is time to use the concept and apply it on the incident data. We first will show the list of cause-asset combinations that led to an accident in the past 11 years. Next, we will explain how to calculate the expected effect of a single cause-asset combination. We end the chapter with the average effects of the relevant incidents.

4.1 Relevant cause and asset combinations

In our mapping of the accidents on the incident process we found that only a few combinations of assets and causes ever resulted in an accident. These combinations are presented in the table below.

Table 2: Accidents per cause/asset combination for the period 1993-2004

ID	Asset	Cause	Personal injury	Damage (=total number)
1	Service line	Excavation	4	47
2	Service line	Own activities	2	3
3	Service line	Soil settlement	1	1
4	Service line	Component defects	2	4
5	Service line	Corrosion	2	9
6	Service line	Other leakage	5	1
7	Mains LP	Other	0	25
8	Mains LP	Component defects	5	6
9	Mains LP	Own activities	18	23
10	Mains LP	Excavation	4	164
11	Mains LP	Soil settlement	0	41
12	Mains LP	Other	2	4

13	Mains LP	Other leakage	1	12
14	Mains HP	Component defects	0	1
15	Mains HP	Corrosion	0	8
16	Mains HP	Excavation	2	68
17	Mains HP	Soil settlement	0	9
18	Gas Meter Installation	Component defects	4	2
19	Gas Meter Installation	Own activities	0	8
		Totals	52	436

Note that the total number of accidents is higher than presented in the previous section. This is due to the extrapolation of the reportable incidents from a 2 year period to an 12½ year period, to match the other accident data.

4.2 Calculating the average incident value

We will demonstrate the calculating principles for the LP mains-excavation combination. This is the combination with the highest number of reported incidents.

Table 3: Average incident value for excavation and LP mains

	Severity levels						
	6	5	4	3	2	1	totaal
Number of incidents	0	0	2	44	94	28	164
Value per incident	1E+08	10000000	1000000	100000	10000	1000	
Value per level	0	0	2000000	4400000	940000	28000	7368000
Average value							44927

In the calculation we determine the number of accidents per severity class. We then multiply the number of incidents per class with the value of an incident in that class and sum those results over the levels, to get the total value of this asset- cause combination. If we divide this total by the number of accidents, we get an average value per reportable incident.

However, this is not the total number of incidents on this asset-cause combination, as most of the incidents do not qualify as reportable. However, for the non-reportable incidents, the available data spans an even shorter period. The incident data is recorded in Nestor Gas from Kiwa Gastec Technology, but it only started in 2003. We took all incidents relating to the combination Mains LP – excavation (some filtering within the database) and extrapolated those results to match a twelve and a half year period. For this, we assumed that the quality of the grid and thus the number of incidents did not change significantly. The results are shown in the table below.

Table 4: Accident probability for incidents

			Incident numbers		Reportable incidents	Probability
ID	Asset	Cause	2004	1993-2004	12½ year period	
10	Mains LP	Excavation	781	9763	164	0.0168

Finally, we have combined the average value and the incident probability into a incident risk figure.

Table 5: Average incident risk

ID	Asset	Cause	Incident value	Incident probability	Incident risk
10	Mains LP	Excavation	44927	0.0168	755

4.3 Results for all combinations

In the table below we show the results for all combinations

Table 6: Risk figures for all cause/asset combinations

ID	Asset	Cause	Value/inc	Probability/inc	Risk/inc
1	Service line	Excavation	172,660	0.00158	272
2	Service line	Own activities	1,066,667	0.00049	527
3	Service line	Soil settlement	1,010,000	0.00006	62
4	Service line	Component defects	3,052,500	0.00178	5,427
5	Service line	Corrosion	1,334,222	0.00104	1,394
6	Service line	Other leakage	60,000	0,00008	5
7	Mains LP	Other	71,200	0.00741	527
8	Mains LP	Component defects	2,538,333	0.00828	21,007
9	Mains LP	Own activities	701,435	0.01172	8,221
10	Mains LP	Excavation	44927	0.01680	755
11	Mains LP	Soil settlement	8,244	0.00576	48
12	Mains LP	Other	282,750	0.00041	116
13	Mains LP	Other leakage	102,333	0.00173	177
14	Mains HP	Component defects	1,000	0.01600	16
15	Mains HP	Corrosion	1,000	0.02286	23
16	Mains HP	Excavation works	51,721	0.15543	8,039
17	Mains HP	Soil settlement	100,000	0.01469	1,469
18	Gas Meter Installation	Component defects	2,050,500	0.00001	23
19	Gas Meter Installation	Own activities	100000	0.00137	137

Multiplying the number of incidents with the associated risk of the incidents and summing over all incident types gives the total safety risk. This is shown in the table below.

Table 7: Total risk, connection risk and individual risk

ID	Asset	Cause	Risk/inc	# incidents	Total risk
1	Service line	Excavation	272	2,384	649,200
2	Service line	Own activities	527	486	256,000
3	Service line	Soil settlement	62	1,313	80,800
4	Service line	Component defects	5,427	180	976,800
5	Service line	Corrosion	1,394	689	960,640
6	Service line	Other leakage	5	1,010	4,800
7	Mains LP	Other	527	270	142,400
8	Mains LP	Component defects	21,007	58	1,218,400
9	Mains LP	Own activities	8,221	157	1,290,640
10	Mains LP	Excavation	755	781	589,440

11	Mains LP	Soil settlement	48	569	27,040
12	Mains LP	Other	116	780	90,480
13	Mains LP	Other leakage	177	555	98,240
14	Mains HP	Component defects	16	5	80
15	Mains HP	Corrosion	23	28	640
16	Mains HP	Excavation	8,039	35	281,360
17	Mains HP	Soil settlement	1,469	49	72,000
18	Gas Meter Installation	Component defects	23	14,077	328,080
19	Gas Meter Installation	Own activities	137	467	64,000
	Total				7,131,040
	Per connection	**7,031,000**			**1,014**
	Per inhabitant	**16,000,000**			**0.445**

In this table we see that the risk per connection is about 1 euro per year, or 45 eurocents per inhabitant. In terms of equivalent fatality risk this is about once every 20 million years per inhabitant. Equivalent risk means that any financial damage is translated into a fatality risk. In the last 12½ years only one fatality occurred, so the actual fatality risk was about once every 200 million years. Both figures are well below the once per million year limit for individual fatality risks.

5 CONCLUSIONS

As we have seen, it is possible to predict the safety of a system based on the number of occurrences of minor incidents. However, there is a point, beyond which the benefit of more data is balanced by the diminishing quality of the data. The data we use in our paper are somewhere in between, being incidents that are serious enough to make a phonecall about. Furthermore, we have seen that only a small number of cause/asset combinations ever became dangerous. Monitoring the numbers on those incidents gives a reasonable feel for the safety of the grid. In total, the gas distribution system is very safe. In terms of direct fatalities it is roughly once every 200million years for an individual, in terms of equivalent damage it is about once every 20 million years.

However, we have also seen that the current data used in the safety indicator is not of a very high quality; Essent scored sometimes almost none of the risks, sometimes almost all. It is more likely this is caused by misinterpretations of data than because of differences in the grid. To make the indicator a really effective regulatory instrument this has to be sorted.

6 REFERENCES

[1] Akkeren, R. v., Y.C. Wijnia, AsM-PR-veiligheidsindicator Gas (AN)-c-I.6.doc; Essent Netwerk B.V.; 2005 (Dutch)

[2] Baarsma, B.E., P.H.G. Berkhout & J.P. Hop (SEO), Op prijs gesteld, maar ook op kwaliteit: de prijs van stroomonderbrekingen- op zoek naar phi, Amsterdam, 2004 (Dutch)

[3] Clark, P.S., Inquiry into national workers' compensation and OH&S frameworks;; Workplace Safety and Health Management; 2004

[4] Hermkens, R., Veiligheidsindicator Gasdistributienetwerken – Implementatie van de veiligheidsindicator, , Kiwa-Gastec Technology, 2005 (Dutch)

[5] Kahneman, D. and A. Tversky, "Prospect theory, an analysis of decision under risk," Econometrica, vol. 47, pp. 263-289, 1979.

[6] Klinke, A. and O. Renn, "A new approach to risk evaluation and management: Risk Based, precaution based and discourse based strategies," Risk analysis, vol. 22, pp. 1071-1094, 2002.

[7] Körvers, P. M. W., Accident Precursors; thesis ; TUE, 2004

[8] Morgan, M. G., H. K. Florig, M. L. Dekay, and P. Fischbeck, "Categorizing risk for risk ranking," Risk analysis, vol. 20, pp. 49-58, 2000.

[9] National Research Counsil (Committee on Risk Categorization), Understanding Risk: Informing Decisions in a Democratic Society. Washington D.C.: National Academy Press, 1996.

[10] Nichols, D.L., Mishap Analysis;; Air University Review, 1973

[11] Perrow, C., Normal accidents: Living with high risk technologies. Princeton: Princeton University Press, 1984.

[12] Saldaña, M. A. M. et al Assessing Definitions and Concepts Within the Safety Profession;.; The international Electronic Journal of Health Education, 2003; 6:1-9

[13] Slovic, P. and E. U. Weber, "Perceptions of Risk posed by extreme events," presented at Risk management Strategies in an uncertain world, New York, 2002.

[14] Whiting, J.F., Risk Tolerability Framework – Developing and implementing a practical workable framework for your workplace; paper presented at APOSHO, 2001

[15] Wijnia, Y.C & P.M. Herder, Modeling interdependency in electricity infrastructure risk, , paper presented at CZAEE 2004

[16] Wright et al, Accident versus near-miss causation;.; Internet: http://www.nae.edu/NAE/engecocom.nsf/754c87f163f599e85256cca00588f49/85256cfb004759c185256dd60053b88d/$FILE/vanderschaaf.PDF

Acknowledgments

The research of the author, described in this paper, is sponsored by the "Next Generation Infrastructures" research programme (www.nginfra.nl).

The safety indicator would not have been possible without the cooperation of many people within Essent Netwerk, Kiwa Gastec Technology, NUON, ENECO and Delta.

ASSESSMENT OF ROAD SIGNS FOR RETROREFLECTIVITY

Stephen J. Howe [a]

[a] *Stephen J. Howe, Senior Manager Infrastructure Operations, Latrobe City Council, P.O. Box 264, Morwell,Victoria 3840 Australia.*

Abstract: Road Authorities in Australia provide a range of road signs, the type, size, placement and manufacture of which is covered by a range of Australian Standards, Design Manuals, and Codes of Practice associated with legislation. However the in-service performance is not currently mandated. Within their Road Asset Management plans, road authorities must describe their road-related services and intervention levels for service provision, as the basis for budgeting, community acceptance and defence of liability claims. Signs must remain visible and legible at distances assumed by design engineers when initially installed, by day and by night. The quantification of in-service night-time visibility and legibility standards applicable to individual signs or even classes of signs is particularly difficult, with multiple, varying factors to take into account in any given circumstance. However, research has determined rules of thumb that can be applied (albeit simplistically) as *de facto* performance standards, and in-service performance can be assessed against the resulting criteria. A key component of the assessment process is determining a sign's 'retroreflectivity' (or reflective luminance). The issues associated with measuring retroreflectivity against objective service standards is discussed in detail, as well as presentation of a regional municipal road authority's recently adopted approach.

Key Words: Signs, Retroreflectivity, assessment, standards.

1 BACKGROUND

Road signs perform a wide variety of tasks on the road network – from guidance on tourist destinations through to statutory control. Regulatory, warning and hazard signs in particular are crucial parts of the designed road safety system, working in conjunction with delineation (centreline and edge marking), road geometry, public lighting and traffic signals to form the overall system of control within which vehicle drivers, cyclists and pedestrians operate.

A large body of research exists which has led to the various state or federal documents that advise road designers on what type of sign to place and where. These typically account for the message or outcome required, vehicle speed, environment (including weather, terrain and complexity of infrastructure) human response factors such as reaction time, and vehicle response factors such as braking distance.

There is an implied assumption in the standard texts that the signs installed will remain fit for purpose throughout their service life, and will be replaced when that service life has expired. However, other than for catastrophic failure (such as occurs with vehicle impacts or extreme acts of vandalism), quantifying when a sign is no longer fit for purpose can be problematic. Standard condition survey approaches do not typically account for the complex design intent, and are often conducted statically, with an observer up close to the sign, i.e. not in a vehicle, at the design speed. In addition to daytime visual condition, the night-time performance of a sign must be considered and evaluated objectively.

Night-time performance is, if anything, more critical, as it is subject to a range of complicating factors not evident in the daytime. The main physical property that contributes to a sign performing at night is its "retroreflectivity". This relates to its degree of luminance when subjected to an external light source, and is a property which can be measured objectively. The sheeting from which Australian road signage is made is manufactured to the Australian/New Zealand Standard AS/NZS 1906.1 "*Retroreflective materials and devices for road traffic control purposes – part 1: Retroreflective sheeting*" (with road edge and centre delineators and high visibility clothing manufactured to other standards in the same series), with different sheeting grades and colours having different associated retroreflectivity properties. The standard is currently under review, with a review draft being made available for comment in late 2005.

2 STANDARDS – COMPULSORY OR VOLUNTARY?

Since the much-discussed 2001 High Court repeal of the so-called "highway immunity"defence against public liability claims, the risk management context in which this subject is set is one whereby Australian road authorities have had to put in place more proactive and better documented road patrol and response systems, set within the context of formally adopted Road

Asset Management Plans – essentially public domain documents. Such plans must document what assets road authorities have, and to what standards they will maintain them.

However, a road authority may still remain open to public liability claims if it can be shown that its standards or systems are 'unreasonable'. Reasonableness will be determined by a Court comparing a 'defendant' road authority's standards with those of its counterparts, and/or with adopted industry standards (as expressed by peak body guidelines, specifications, Codes of Practice and the like). Few standards are mandated by law, and in that sense, road systems differ markedly from say, Buildings, in terms of mandatory minimum standards of design, construction and maintenance.

Road signs are a good case in point. Whilst *temporary* road works signs in Victoria have been prescribed as to the applicable AS/NZS 1906.1 sheeting class since mid 2005 (under Codes of Practice subordinate to the Road Management Act 2004 and its associated Regulations), *fixed* road signs have not. And even the temporary road works signs, which are required by the Victorian Code to be "*constructed of retroreflective material meeting or exceeding the performance standard for Class 1 material*", do not have their long term performance described or even discussed by such legislation…just their initial 'construction' (i.e. installation).

It is impractical and unsustainable (as well as unjustified by any research) to simply require all signs to perform to the 'as new' retroreflectivity standard of a given class of sign sheeting. The problem is that <u>no in-service retroreflectivity performance standard exists</u>. Appendix J of the current standard contains decay curves for most (but not all) sheeting types and colours, but the Appendix is only "advisory", and therefore does not even form part of the notional official framework provided by the standard. And even with the decay curves available (which are based on Jenkins and Gennaoui, [1]), the threshold for minimum performance is not defined, making the road asset managers' task seemingly impossible. How one Victorian municipal road authority is dealing with this dilemma is discussed further later in the paper.

In the United States of America, considerable effort has been focussed on this issue by peak industry and research bodies as well as regulators, for a number of years. In 1991, the Intermodal Surface Transportation Efficiency Act required "*the Secretary of Transportation…*(to)*…revise the manual of Uniform Traffic Control devices to include…a standard for a minimum level of retroreflectivity that must be maintained for pavement markings and signs, which shall apply to all roads open to public travel*" (as cited in Kalchbrenner, [2]). This focus is subtly, but importantly different to the current Australian approach - which is only about the standard of initial manufacture, and (except in the case of workzone temporary signs), not even mandated by legislation anyway.

Since this legislative imprematur, the relevant bodies in the U.S.A. have been busy, with the Federal Highway Administration issuing minimum retroreflectivity guidelines in 1997, and updating them in 2003 (see Carlson and Hawkins, [3]). Road authorities and road-research bodies have trialled various technologies in their quest to comply with the resulting (mandated) industry standards, and as a result the industry is more rigorous and mature in its approach than the average Australian counterpart.

However, momentum would appear to be building in Australia for mandated minimum standards, as the 2004 Inquiry into National Road Safety [4] included the following recommendation – "*that design and maintenance standards on the national highway system conform with world's best practice*", and a statement that "*minimum design and maintenance standards should be created and enforced*". Though these statements appear limited in intent to state and federal highways, it could (via the 'reasonableness' test Courts will apply) ultimately be broadened to include <u>all</u> public roads over time. The issue of what standards could 'reasonably' be applied is discussed further below, in the conrtext of having first considered the multiplicity of issues that add complexity to the matter.

3 ISSUES ASSOCIATED WITH SIGN PERFORMANCE (INCLUDING RETROREFLECTIVITY)

There are multiple, overlapping factors that apply to the issue of in-service performance of road signs, with retroreflectivity being merely one (albeit an important one). Some relate equally to day or night situations, while others are more critical only at night. These include:

1. **"Visual Clutter**". Signs are only one of the many visual signals being interpreted by drivers, which they need to process in order to react appropriately. In addition to the environmental and driver capacity factors discussed below, the level of visual clutter is a factor that significantly affects whether a sign can practically communicate its message. Other signs, advertising billboards (some of which can be very attention grabbing!), traffic signals, linemarking and the proximity and behaviour of other traffic all contribute to the visual clutter phenomenon. Adams et al [5] found when testing drivers in a driving simulator with eye-tracking capacity that "*the density of traffic has a profound effect on their behaviour in relation to signs….subjects who were exposed to more traffic were more likely to have accidents, were more likely to exceed the speed limit, and were less likely to stop at traffic lights, thus demonstrating what can only be described as a frustration effect*".
2. **Vehicles**. These can vary enormously - headlights can be mounted at different heights on vehicles, and the distance between headlamp height and driver eye height also varies. These affect the degree of sign luminance perceived by the driver (e.g. compare a sedan with a truck), as light is reflected within a limited 'cone' back towards the light source. Also, due to the changes in vehicle headlamp manufacture that occur over time and on different continents, the quality of light

being delivered for reflection can vary with the date and country of vehicle manufacture, making the concept of the "standard design vehicle" hard to sustain. Fortunately, it would appear that recently developed headlamps - which use gas discharge instead of tungsten filament or quartz halogen technology - whilst adversely affecting the reflected luminance of *certain* sign colours, have not made a significant difference overall, with Venkatachalam and Smith [6] concluding that "*new gas discharge lamps will not decrease the visibility of coloured road signs, compared to conventional tungsten halogen lamp*". However, in contrast to the quality of light, the strength of the the light beam can vary by as much as 25%-300% depending on headlight type and whether high or low beam is in use, just within the vehicles manufactured in the U.S.A.

3. **Environmental conditions** – issues such as rain, fog, and topography will affect a signs effectiveness, both directly (in terms of reduced visibility) and indirectly (in terms of diverting the driver's concentration from the sign recognition task). Topography is important to account for when first installing signs, as vehicle headlights will tend to over or undershoot a sign (until closer than the equivalent distance on flat terrain), depending on whether going up hill or down, and whether or not the curve of terrain is concave or convex. Excessively winding roads can have a similar 'masking' effect. These factors point to careful use of 'standard' height mounting poles (as one size may not fit all), and advance warning signs. The issues of rain and fog are even more complex. Rainfall performance is at least nominally covered by a 30% maximum allowable degradation in performance allowed in newly manufactured sheeting (as per the Standard), but whether this is adequate has not been proven, and in any case every rain storm will differ in intensity and direction. The issue of performance in fog has been studied in recently in Japan (Munehiro et al, [7]), but although it was determined that daytime fog has a greater effect on sign visibility than night time fog, no specific minimum performance criteria has emerged as yet. It is perhaps best to remember that road rules require that drivers adjust their speed to account for all relevant conditions, and that road engineers cannot be expected to cover all contingencies.

4. **Ambient light levels**. While street lighting is helpful for some aspects of navigating (e.g. being able to perceive general surrounds without being limited to the vehicle headlights capacity to light up the vicinity), it can actually hinder sign visibility, by reducing the visual impact of the reflected sign luminance on the driver. Sign sheeting is not manufactured to intentionally reflect anything other than headlight luminance, so the degree to which the street lighting itself 'lights up' the sign may not compensate for the reduced impact of sign luminance on driver perceptions. It also appears that the dark 'surround' provided by low levels of ambient lighting assists with the overall conspicuity of the sign. Similarly, the prescence of oncoming traffic (and whether or not they have high or low beam on) complicates the issue even further, and <u>always</u> at the detriment of sign visibility. Conversely though, in certain situations it may be necessary to actually illuminate signs intentionally, as pointed out by the U.S.A.'s Federal Highway Administration [8], "*auxillary illumination may be required on curved roads, areas of high and spotted illumination, conditions of frequent fog and dew, and at action situations such as at highway exits*". This compensatory measure recognises the inability of vehicle headlights to deal with extreme situations.

5. **Driver capacity factors**. Drivers can have a reduced capacity to respond to signs for a range of reasons – tiredness, drugs/alcohol, distraction on other matters (including stress and mobile phone use) and physical limitations such as colour blindness or extremes of light sensitivity. However the single most relevant factor studied is driver age. Advocates for older drivers (such as the Council for the Ageing, and various automobile associations, which often have a high proportion of older members) often point to the crash statistics on younger drivers, and the fact that the issues that face older drivers (such as hard-to-read signs) are issues for <u>all</u> drivers. These are valid points, but fail to recognise the uncomfortable realities revealed in the results of actual studies:

    i. Chrysler et al [9] commented that "*older drivers, in particular, often have trouble reading street name signs because of poor sign design and because of vision problems associated with ageing*", and found that drivers with a mean age of 71 needed brighter signs than other drivers in order to be able to read signs in complex traffic situations.

    ii. Pletan et al [10] have commented that "*a 70 year old person needs about four times more light to achieve the same visual acuity as a 25-year-old*".

    iii. Sivak and Olsen [11] found that "*the usually observed age-related performance decrement on nighttime legibility tasks is the result of visual-acuity deficits*". Whilst some of the deficits may be correctable with increasingly powerful glasses, this relies on the awareness of the aging driver of the issue and their willingness to act on that awareness. Whilst some older drivers exhibit compensatory behaviours (driving less, driving on known routes only, driving slower), this may not always be sufficient. King (2000) found that "*at some point, loss of judgement means that the compensatory efforts are no longer effective, and may even be counterproductive*".

    iv. King [12] also confirmed from a very detailed study of crashes and age, the simple fact that "*crash involvement …increases steadily with age*", and found that, in addition to visual acuity defects, that a jump in rural accident rates for drivers aged 80 and over may "*be due to a threshold of cognitive impairment being reached*".

6. **Sign orientation**. If not installed at the correct angle to the road, a sign will fail to reflect light at the angle assumed by sheeting manufacturers and traffic engineers, causing reduced or in some cases even excessive visibility. One (unsubstantiated) submission to the Australian Inquiry into National Road Safety [4] stated "*this is a management item that appears in all roads manuals yet in Victoria it is estimated that no better than 10% of signs are installed at the*

correct angle to avoid this "white out" problem…there is a plethora of signs (possibly as many as 80%) that are installed square to the road".

7. **Sign type** – it seems that, all other things being equal, drivers notice certain signs more than others. Charlton [13] reports in a study on hazard warning signs that "*of the signs tested, road works and school warning signs were most often detected, remembered and understood. Slippery surface warnings were associated with some of the lowest detection and comprehension rates*". Other studies have shown that the familiarity of the driver with the surroundings is also a factor. It seems that drivers almost 'don't want' to register certain signs as being important, particularly when repetition of exposure has occurred.

8. **Sign Cleanliness**. As road grime builds up on a sign, its daytime and nighttime performance invariably suffers. Jenkins and Gennaoui [1] cite that in regards to retroreflectivity, "*washing produced a 10-20% improvement, and could be much higher for particular sites that caused greater pollution*." Conversely, aggressive washing to remove grime or graffiti (using certain chemicals or abrasive rubbing) was found to reduce retroreflectivity. Sheeting manufacturers should be consulted before chemicals are used on particular contaminants resisting gentle cleaning methods.

9. **Sign retroreflectivity**. This is a physical property of the reflective sheeting, screen printing or multiple sheetings that make up a sign, and is dependant on the class designation to which it is manufactured (class 1, 1A, 2 etc), colour, and age. Extensive studies have documented the age-related decay that occurs in this material property in-service, which is slightly worse in industrial environments or those with extreme ultraviolet radiation levels (e.g. northern Queensland). Most real signs are made up of more than one colour, so this issue is in fact more complicated than it first appears. Not only must the *individual* colours have a sufficient level of luminance for their service life, but the *ratios* of luminance must be sufficiently different over the same period in order for the sign to remain effective. **Note**: the Australian Standard actually specifies minimum luminance contrast ratios, but fails to give comprehensive guidance, being based on research that did not account for the full range of colour combinations present in real signs.

The latter issue creates its own problems, as signs are visible as a function of their overall sign luminance, and legible as a function of the individual colour luminance levels and ratio – an issue discussed further below. Determining overall luminance requires the use of either detailed calculations based on weighted averages of area per colour, or simplistic approaches such as averaging the scores of the individual colours (which can be under or overconservative by a large degree). The test methods in the standard only really address the process of determining the retroreflectivity score for an individual sheeting colour, whereas in real life, assessors have to deal with (and decide the 'fate' of) an entire, multicoloured sign.

In the absence of performing detailed calculations or using results of specific research, the only 'safe' approach appears to be to assess the individual colours against a standard, and assess the luminance contrast ratio. Overly simplistic approaches, such as attempting to set a single standard level of retroreflectivity for a given sign type therefore currently appear to defy achievement, and should not be the expectation of practitioners.

Researchers typically talk about several aspects of 'seeing' a sign (and other traffic control devices). Firstly there is visibility – i.e. "can a sign be seen?" Then there is "conspicuity" - defined by the International Commission on Illumination as "*the attribute of an object within a visual context which ensures its prescence is noticed at the pre-attentive level of processing*". Put simply, a sign may be visible, but not 'conspicuous enough' to be noticed in certain contexts. Lastly, there is legibility. This is not so important for signs that communicate their message largely by their shape and colour, but critically important for signs that have to be read in order to serve their purpose. How drivers acquire and use all these cues is a very complex science, with different aspects being sensed and processed by different cognitive processes. Even legibility has both a static and dynamic aspect (as discussed by Schieber, [14]).

It should be understood that drivers progress through phases of 'seeing' a sign, some of which are subliminal, including recognition of a signs' shape and colour even before the written message becomes legible. Where there are too many competing or incapacitating factors, invariably a sign is not 'seen' at all, and is (not surprisingly) ignored, sometimes at the cost of an accident. Recent research using in-vehicle eye-tracking systems (Schieber and Burns, [15]) reports that "*reading a mission critical highway sign requires more attentional resources than expected*".

Jenkins and Gennaoui [1] summed up the core issues to be managed by road authorities as follows – "Locating the sign is likely to be more influenced by the sign luminance. Legibility of the sign will be influenced by internal contrast and luminance". However, they also noted that managing these issues is subject to an example of the law of diminishing returns; "a large change in luminance will result in a small change in legibility distance. This is shown explicitly in a paper by Colomb and Michaut (1986), where an increase in luminance by a factor of 3 results in a 15% increase in legibility distance".

It should not be considered appropriate for road management authorities to "punish" the society which funds its road infrastructure by assuming all possible worst-case scenario's should apply when designing the solution to a signage problem, as the resulting cost would be unaffordable and put no onus back on the individual. Thus it is not necessary to assume - as some researchers have done - that road signs must be designed and maintained so as to cater for older drivers, who have consumed alcohol, are driving in the rain, and have poorly maintained vehicles etc.

With regards to the legibility vs. visibility issue, it also seems unnecessary to require signs to be legible at the same distances they need to be visible and/or conspicuous, and it suggested that future research address establishing not only ready-reckoners for minimum sign retroreflectivity and luminance contrast, covering a range of situations (as the U.S. authorities have established), but also distance standards that differentiate between the legibility and visibility distance criteria, especially as these will have considerable impact on the resulting required retroreflectivity levels. In the meantime, the landmark Australian study by Jenkins and Gennaoui [1] gave numerous examples of sign scenarios over a range of speed limits and their resulting (dynamic) legibility distances were typically in the 40-70 meter range. This of course assumes that the engineering designer will use the appropriate size sign and lettering for the situation, chosen from predetermined standard templates.

However, it is important to note that much of the early research combined the use of *static* legibility distances (sometimes gained in laboratories rather than in the field) with research on human and vehicle response times to develop models for dynamic (at speed) sign legibility. It is only very recently that dynamic legibility has been able to be tested in the field, and even then only for a limited range of signs, at one speed (60 mph). Schieber [14], reporting on the findings of this research, found that dynamic (night time) legibility indexes declined below static distances by around 30%, falling even below the required standard. The study also found that dynamic night time legibility fell by around 12% for an 85% drop in sign retroreflectivity, which corresponded to the "*new end-of-service-life reflective minimums recently proposed by the Federal Highway Administration*", and commented that "*under most conditions, a life cycle reduction in performance of this magnitude would probably be acceptable*". If true, this would indicate that retroreflectivity is less important a factor for true driver sign legibility (though not visibility and conspicuity) than sign and lettering size, and may lead to a need in future to upgrade the standard approaches to sign sizing. Clearly more dynamic, field test-based research would be required in order to justify such a change and the resulting cost increases.

It is also noteworthy that the retroreflectivity minimums *seemingly* allowed by the U.S. authorities (presumably for all sign needs, not just legibility) fall way below both the warranty standards being offered by sheeting manufacturers in Australia and the 'as new' levels required by the standard. Road authorities will therefore need to decide whether the residual retroreflectivity standards they adopt and enforce in specifications are the mandated minimums (if set, which has yet to occur in Australia) or the considerably higher sheeting warranty levels. They could of course adopt both, which infers keeping signs for considerably longer than the warranty period so as to maximise useful life and minimise life cycle cost.

4 APPROACHES TO TESTING FOR RETROREFLECTIVITY

Figure 1 – Portable Retroreflectometer

There are various ways of testing for retroreflectivity. The crudest, and most time consuming, is not to carry out a retroreflectivity test at all, but simply to carry out night time sign audits, whereby all signs are checked by a driver and the visibility and legibility distances measured and compared with specified distances – which should be those expected by design criteria relevant to the situation and design speed. Whilst fraught with an unavoidable degree of variability (in terms of the number of factors that can vary in a given measurement situation, including differences between observers), this method is at least compatible with real life. Results gained from scientific methods (though more objective) have to be seen in the broader context of the wide range of factors that determine actual sign performance anyway, *many of which are unrelated to the sign's observable physical properties*. Unless all sign testers are required to be accredited road safety auditors (which would in most cases not be feasible for cost reasons), approximate methods using 'reasonable but not excessive' levels of conservatism should be seen as appropriate.

The most commonly available scientific method (though not in apparent use in Australia) is the use of a portable Retroreflectometer, also known colloquially as a 'Photometer', shown in Figure 1 above.

The studies that led to the current Australian Standard were based on the results of using these devices in the field across Australia. They are available with time saving attachments for reaching high-up signs, barcode scanning and GPS integration. They are commercially available via import and can be used to test signs according to the Australian or ASTM standard ASTM E 1709 – 00 [16]. The test methods vary little qualitatively, principally on the number of sample readings taken (the American standard specifies four readings per colour, whereas the Australian makes no mention of number of readings). The observation angle (α) in the American standard is the reverse of that in the Australian Standard [17], but that can be presumably be attributed to the fact that most road signs in the U.S.A. are on the right hand road side, not the left (as that is the side of the road Americans drive on).

The European standards (EN 12899-1/DIN 67 520) use wider illumination and observation angles in tests, so equipment configured for those standards should not be used unless they are also adjustable to the AS/NZS or ASTM methods. An example of a portable retroreflectometer in use can be seen in Figure 2 below:

Once a sign is tested, its retroreflectivity score can be compared with a designated standard, so as to evaluate the need for replacement. It is unrealistic to apply the 'as new' standard, and not justified by any studies. Alternatives could include the performance standard contained within manufacturers warranties, which are typically 80% of the 'as new' standard after 10 years for class 1 sheeting, or 50% of 'as new' standard for class 2 sheeting after 7 years, something the draft revision to AS/NZS 1906.1 (1993) alludes to. This contains an in-built margin of safety, as the 'as new' retroreflectivity levels which sheeting manufacturers are obligated to provide if required by clients to manufacture to the standard are themselves lower than results typically gained from field observations on new signs. For example, if a class 1 white sheeting material is to meet the standard, it must have a Coefficient of Luminous Intensity (CIL/m2) of 250 cd/lx.m2, yet the decay curve formula has been found by research to be:

Figure 2 – Portable Retroreflectometer in field use

CIL/m2 = 294 – 4.1 y (where 'y' is time in years).

Thus a new class 1 sign placed in the field typically has an average CIL/m2 of 294 cd/lx.m2, a providing an initial 17.6% margin of safety, even before the sign starts dropping below the 'as new' standard. It is possible of course for a sheeting manufacturer to produce sheeting which only just meets the standards, in which case the usual safety margin will not exist - though as noted above, in the U.S.A., the in-service minimum values allowed appear to be considerably lower than the as-new standard anyway.

It is important to note that whatever the standard chosen as the notional replacement threshold, it may be prudent to carry out night time cross-checks on all signs with retroreflectivity scores either just above or just below the intervention threshold, so as to ensure only those signs that are genuinely not performing are replaced. It must be remembered that retroreflectivity scores are themselves only an *indicator* of luminance, and it is actual luminance and its value to a real driver in a real traffic situation that ultimately counts.

Some road authorities may even conservatively decide not to even test signs at all, but simply replace them when they reach either a predetermined age, or the warranty date, though these apporaches require either date stamping to exist on all signs (rare) or comprehensive records on when individual signs were installed (rarer). Signs are generally meant to be date stamped, but errors in manufacturing and quality checking mean that most road authorities will have a large number of signs for which they have no idea regarding sign age. Thus they will need to either qualitatively test such signs after all, or simply act conservatively and replace them (possibly needlessly). Whilst there is a cost associated with performance based in-service testing, at least it takes a 'just in time maintenance" approach, and is not overconservative. Studies on which method provides the lowest life cycle cost do not appear to have been done yet.

In the United States, where imposition of standards and concerted research efforts have created a dual impetus for best practice, various alternate approaches have been tried. Maerz [18] reports on a prototype digital video image analysis system in Missouri, stating "*the results show the feasibility of developing a mobile vision-based system to classify and measure the visibility of road signs. The results also showed a rather poor correlation between retroreflectivity and visibility. Retroreflectivity was found to be a poor predictor of visibility of white, yellow and..orange signs…a relatively good predictor of the visibility of red, and to a lesser extent of green and blue signs. Brown signs were found to be of low retroreflectivity and visibility*". Maerz also claimed that "*the method …is the closest possible analog to what the eye sees when looking at signs under the normal illumination provided by the headlights*", but also reported limitations, such as "*the method should be used at night and may be limited to use with high beams*". The latter would be problematic, as road rules typically require drivers to dip headlights when in the prescence of oncoming traffic.

Long [19], cites another, similar technology being used at the Michigan Department of Transportation, the Mobile Evaluation of Traffic Signs (METS) project, which combines video cameras, a calibrated light source, laser range finder and computers, on board a specially configured van. The METS system is a good high-volume, safe method, but has the shortcomings associated with all approaches based solely on retroreflectivity, in that the relationship to actual luminance and thus sign visibility is only an *implied* relationship, without a direct correlation that takes in the many minor variations in sites and signs. The very recent work by Schieber [14] is similar, and also combines in-vehicle digital video capture and GPS, but is not a sign replacement assessment methodology, rather being aimed at testing research assumptions about legibility of signs.

The approaches to standards and testing for sign retroreflectivity of the various Australian state road authorities (based mainly on internet-based research) appear to vary considerably. Whilst it was normal practice for such authorities to have initial manufacture supply specifications that link to the current Australian Standard, few seemed to address the residual performance issue:

- In Victoria, the approach seems to be to use an older driver (55 plus) and visually assess visibility at night, apparently at a standardised distance (though this does not appear to address the issue of design speed affecting the required visibility distance). Oddly enough, by way of contrast, detailed test procedures using portable retroreflectometers are used for pavement markings;
- The approach in the Northern Territory, New South Wales also seems limited at present to testing pavement markings using portable retroreflectometers. Procedures on testing of signs was not able to be located in lists of test procedures related to infrastructure maintenance ;
- In Queensland, sign maintenance procedures discuss both night time visual assessments and use of retroreflectometers, without giving further specifics or residual performance standards;
- In South Australia, the only references discoverable related to warranties on residual performance in supply specifications.
- In Western Australia, there was both expected warranty performance, and a table specifying minimum retroreflectivity values per colour for different signs, and minimum luminance contrast ratios, maximum luminance factors for certain colours and a reference to the Australian Standards test procedures.

The Western Australia approach was therefore the most sophisticated approach able to be found. The Luminance contrast ratios listed were lower than those in Appendix B of the standard, but higher than the absolute minimum of one stated by Jenkins and Gennaoui [1] and coincided with the recommended limit suggested by the same authors. The minimum retroreflectivity scores were far lower than the 'as new' and warranty values in some cases, and therefore would appear to borrow heavily from the approach being taken by U.S.A. highway authorities. The maximum luminance factors stated were not justified, as the Australian Standard does not in fact set maximums for the specific colours listed.

In regard to those authorities requiring performance warranties from sign suppliers, it was not demonstrated how such warranties would ever be qualitatively tested - if there is in fact no measures put in place to either test all signs or at least those signs being replaced for perceived non-performance. In such cases the sign sheeting suppliers will never have to practically honour their warranty statements, which defeats the purpose of requiring them in the first place in a contract.

In summary, the Australian approaches to this issue – as evidenced by the State Road authorities published frameworks - appear to be inconsistent and lacking a genuine 'in-service' performance standard in almost all cases. There also appears to be no evidence of sophisticated, automated (typically vehicle-based) approaches to performance testing being used, such as the METS or digital video analysis techniques used in the U.S.A. This is odd, as such advanced technologies are being used in Australia for other aspects of road infrastructure assessment, such as road pavement condition rating.

One issue which has emerged recently (and could cause problems for authorities seeking to use retroreflectometer-based approaches) is that of the newer 'flourescent' sign colours. The draft revised standard includes performance information on the manufacture of the flourescent colours which have come into use in recent years, yet portable retroreflectometers may not configured to test for them, if they are configured and calibrated to assess standard colours only. In such cases, night time visual assessments against acceptable visibility/legibility distance standards may be the only means of testing performance.

5 COMPARISON OF APPROACHES, COMMENTARY AND RECOMMENDATIONS

In terms of testing cost alone, of the two *scientific* methods known to the author to have a quantified cost, use of a retroreflectometer is more cost-effective. In the initial trial project at Latrobe City (a regional municipal Council in Victoria), which included the initial purchase of a Trimble datalogger/GPS and related software, the Retroreflectometer purchase, and labour over a three month period, the cost per sign was approximately $2.70-$3.30AUD. This cost assumes the amortising of start-up capital costs over 3 years, and excludes the cost of transport, as the assessor either travelled on patrol maintenance trucks which were doing regular cyclic inspections (in rural areas), or used existing organisational fleet cars for township-based data collection. This cost also included capture of the initial management data for the signs, not merely the condition/retroreflectivity assessment, so in fact would go down slightly for future inspections focussing solely on sign retroreflectivity. The actual cost for another authority would of course vary, depending on the number and density of signs managed by it, and whether or not they already possessed portable data entry/GPS devices.

This cost compares well with the "METS" method cost per sign of around $4.55 AUD (which also excludes transport costs, and uses a three year capital amortisation), which would be higher if CPI effects were taken into account since the METS approach was first reported….though it is also possible that the associated technologies may have matured and/or reduced in price since METS was first implemented.

The cost of applying other testing/replacement methods (such as simple installation date-based replacement, or night time subjective visual assessment) is unknown, as no published data is available - though could no doubt be available from those road authorities which use such approaches. As discussed earlier, the approach with the lowest life cycle cost is as yet unknown, as a degree of overconservatism may exist with the less scientific methods, that results in unnecessary sign replacements and thus a higher cost - that may or may not exceed the savings made in not using in-field performance assessment methods.

Given the complexity of the issues and likely variance in capacity of different road authorities, it is perhaps worthwhile to note the comment made by Carlson and Hawkins [3], "*Agencies should have the option to select the methods or combination of methods best suited to their needs and resources*", and "*should have a 6-year period to implement methods*". What should <u>not</u> be acceptable is for road authorities to simply ignore the issue and do nothing – a crude approach is after all better than no approach at all. It is the author's opinion that many road authorities in Australia have yet to grapple with this issue in any comprehensive way, and are probably exposed in terms of public liability, given that signs play a crucial role in road safety. Compared to the many bodies actively researching and implementing rigorous approaches in the U.S.A, Australia would appear to be well behind in terms of world's best practice.

Latrobe City's emerging approach is one of combining the K.I.S.S. principle with a reasonable degree of sophistication, and has been based on the lessons learnt during an initial three month trial, and attempts to deal with the complexities associated with real signs and the issue of balancing risk with cost, as well as the errors of omission of the past. The overall (recommended) approach is as follows:

- All signs to be captured initially, including GIS location. Regulatory signs also assessed for Retroreflectivity and management data (e.g. type, date stamp etc);
- Set minimum visibility and (separate) legibility distances for signs, recognising differences of road speed and location (rural vs. urban), and the results of emerging research on dynamic sign legibility;
- Set minimum retroreflectivity and luminance contrast ratios for common sign colours;
- Give highest priority – in terms of inspection effort - to <u>safety</u> related signs: Regulatory (other than parking control), Temporary Works, Warning and Hazard signs;
- Give second priority to street name blades (of importance to both residents and emergency services)
- Give lowest priority to tourist, guide/advisory and parking control signs;
- Assess all signs on a cyclic basis, with higher priority signs inspected more frequently;
- Assess all high priority signs that are close to 'intervention level' (i.e. replacement) annually, once they are close to the warranty limit (if date of manufacture known) or the luminance-based performance limits (where no date known);
- Only clean signs if found to be below required performance levels when tested, then test them again. This may lift cleaned signs back into the 'acceptable performance' range (or confirm the need for replacement), and saves needless expenditure on cleaning <u>all</u> signs when there is not sufficient cause;
- Perform the cyclic tests by day – if the signs do not pass the daytime criteria such as daytime visibility distance, there is no need to even test for the night time criteria.
- Night time criteria are assessed via portable retroreflectometer testing. Signs which fail marginally are double checked at night using distance criteria to confirm failure before replacing.

The latter recognises the difficulty in using a single numerical result from a scientific test to justify sign replacement, as the sheer number of complicating factors can mean that a sign may not have 'failed' after all when observed by a real driver at night. In that sense, the approach described above can be described as a 'fuzzy logic' system, in that it "*is a way of reasoning that can cope with uncertain or partial information; a characteristic of human thinking and some expert systems*" [27].

Latrobe City is now applying this procedure and is actively replacing failed signs as a result. Ultimately, when it has sufficient data to be able to correlate the legibility and visibility distances of failed signs with their retroreflectivity scores, it will be able to refine its process further, possibly to the point of dispensing altogether with the night time "cross check".

6 FUTURE DIRECTIONS

There is a general trend towards improving the standard of road infrastructure that is arising from greater levels of public awareness, better research, the trend towards increased litigation against road authorities on public liability grounds, and improved asset management awareness. Unfortunately, this is in a climate whereby many road authorities are faced with a funding gap (see Burns et al, [20]) due to the fact that their networks are approaching the renewal point in their life cycle, and have to operate with budgets that have been based on CPI-based increases rather than on a clear awareness of asset funding needs. This management problem has often been exacerbated by unsustainable decision-making by elected decision makers, who often expect road infrastructure renewal and maintenance to 'just happen', and have been known to favour politically popular new assets and services at the expense of sustaining existing infrastructure.

With ever-evolving intelligent transportation systems, some funding decision makers may argue that there is no great need to invest resources in maintaining signage networks to performance standards. But as Adams et al [5] noted as long ago as 1998, "*despite the advent of in-car information systems…and the dedicated road infrastructure for vehicle safety in Europe programme….road sign fixed warning signs will remain the primary hazard warning system for the forseeable future*".

That being the case, it is suggested that the Federal Government should make it mandatory (probably via State Road legislative frameworks) that minimum construction and maintenance standards are set for ensuring that the various aspects of sign luminance - and thus the resulting qualities of visibility, conspicuity and legibility - of road signs remain at reasonable levels, and that road authorities are required to demonstrate that they actually assess them proactively. However, this does not mean that the Federal Government needs to set the standards – road authorities should be given some leeway to decide their cycles of inspection and method of assessing and setting intervention levels, in keeping with their communities expectations and capacity to pay. A Federally funded project to develop maintenance guidelines (with guidance on the full range of approaches and issues) to establish benchmarks would greatly assist the road management industry, as would a more equitable distribution of road-related fuel taxes.

7 REFERENCES

[1] Jenkins, S.E. and Gennaoui, F.R. (1992), "*Terminal Values of Road Traffic Signs*", Special Report No. 49, Australian Road Research Board.

[2] Kalchbrenner, J.S. (1999), "Signing and Marking Materials", *TR News*, edition 205 (November-December 1999), Transportation Research Board, National Research Council, Washington DC, U.S.A

[3] Carlson, P.J. and Hawkins, H.G. (2003) "Updated Minimum Retroreflectivity Levels for Traffic Signs", Report No. FHWA-RD-03-081, Federal Highway Administration Office of Safety, Washington DC, U.S.A.

[4] "*National Road Safety – Eyes on the Road Ahead*" (final report of the Inquiry into National Road Safety), Federal Parliament Standing Committee on Transport and Regional Services, Canberra, 2004.

[5] Adams, A., Bright, J. and Searle, B. (1998), "Signs of trouble to come? A behavioural assessment of the effectiveness of road warning signs", University of New South Wales/Federal Office of Road Safety.

[6] Venkatachalam, K., and Smith, G. (2000), "Gas discharge headlights and visibility of coloured road signs", *Journal of Clinical and Experimental Optometry*, Australia, September-October 2000.

[7] Munehiro, K., Tokunaga, R. A., Asano, M. and Hagiwara, T. (2005), "Effect of Time and Foggy Conditions on Subjective Visibility: Evaluation of Retroreflective Traffic Control Devices", *Transportation Research Record*, Issue 1991, Transportation Research Board, Washington DC, U.S.A.

[8] "Minimum Retroreflectivity Levels for Overhead Guide Signs and Street Name Signs", Publication No. FHWA-RD-03-082, U.S. Department of Transportation, Federal Highway Administration, 2003.

[9] Chrysler, S., Stackhouse, S., Tranchida, D. and Arthur, E. (2001), "Improving Street Name Sign Legibility for Older Drivers", *Proceedings of the Human Factors and Ergonomics Society* (Report No. R0211201), 2001.

[10] Pletan, R. (Chair, 2000), "The Aging Eye and Implications on the Visual Task of Driving", Report A3C12-02, *Transportation Research E-Circular* No. E-C022, Transportation Research Board, National Research Council, Washington DC, U.S.A, November 2000.

[11] Sivak, M., and Olsen, P.L. (1982), "Nighttime legibility of traffic signs: conditions eliminating the effects of driver age and disability glare", *Accident Analysis and Prevention*, Volume 14, Issue 2, April 1982, Elsevier Publishing, U.S.A.

[12] King, M. (2000), "Intersection crashes involving older drivers: Relative involvement in urban and rural areas at night and by day", in *Road Safety Issues for Older Road Users – Selection of papers from the Older Road User Safety Symposium* (Brisbane , 2000), Queensland University of Technology, January 2004.

[13] Charlton, S.G., "Conspicuity, memorability, comprehension and priming in road hazard warning signs", *Accident Analysis and Prevention*, Volume 37, Issue 6, December 2005, Elsevier Publishing, U.S.A.

[14] Schieber, F. (2005), "Effects of Driving Speed and Sign Reflectance on Night time Highway Sign Legibility Distance and Reading Time Requirements", *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*, Orlando, 26-30 September 2005, U.S.A.

[15] Schieber, F. and Burns, D.M. (2006), "Legibility and Eye Fixation Patterns While Reading Highway Signs Under Dynamic Nighttime Driving Conditions", *Proceedings of the Transport Research Board 85th Annual Meeting*, January 22-26, 2006, Washington DC, U.S.A.

[16] ASTM E 1709 – 00 (2000) "Standard Test Method for Measurement of Retroreflective Signs Using a Portable Retroreflectometer", American Society for Testing Materials, U.S.A., 2000.

[17] AS/NZS 1906.1: 1993 "Retroreflective materials and devices for road traffic control purposes – part 1: Retroreflective materials", Standards Australia.

[18] Maerz, N. (2003) "Automated Mobile Highway Sign Retroreflectivity Measurement", National Cooperative Highway Research Program (IDEA Project 75), National Academies of Science-Transport Research Board, Washington DC, U.S.A.

[19] Long, D. (1997), "Michigan DOT reflects on signs", *TR News*, edition 192 (September-October 1997), Transportation Research Board, National Research Council, Washington DC, U.S.A.

[20] Burns, P., Hope, D. and Roorda, J. (1998), "Facing the Renewal Challenge", Victorian Local Government Infrastructure Study, Office of Local Government, 1998

[21] DR 05585 "Retroreflective materials and devices for road traffic control purposes – part 1: Retroreflective materials" (Draft for Public Comment), Standards Australia, 2005.

[22] VicRoads *Traffic Engineering Manual Volume 2 – Signs and Markings* (3rd edition), December 2001.

[23] "*Road Safety in Rural and Remote areas of Australia*", Austroads Publication No. AP-R273/05, Sydney, 2005.

[24] Jenkins, S.E. (2006), "Performance of Modern Retroreflective Sheeting Material in the Road Environment: Visual Acquisition and Sign Legibility", *Proceedings of Transport Research Board 85th Annual Meeting*, January 22-26, 2006, Washington DC, U.S.A.

[25] Jenkins, S.E. (2006), "Performance of Modern Retroreflective Sheeting Material in the Road Environment: Legibility of Two Classes of Sign Sheeting", *Proceedings of Transport Research Board 85th Annual Meeting*, January 22-26, 2006, Washington DC, U.S.A.

[26] Submission to the Parliamentary Road Safety Committee Inquiry into Older Road Users, Council on the Aging (Victoria), 2002.

[27] www.wiley.com/college/turban/glossary.html

Acknowledgments

The author wishes to acknowledge the following individuals and organisations who contributed advice, assistance and/or documents to the author: Australian Transport Safety Bureau, Frank Schieber (Heimstra Human Factors Laboratories, University of South Dakota), Bob Munns (3M Traffic Safety Systems Division), Lee Yuill.

TOOLS FOR ANALYSING LARGE NUMBERS OF INSPECTION DATA, USED TO ASSESS REMNANT LIFE OF PIPEWORK

F Hoeve S Terpstra P van de Camp & R Paleja

Shell Global Solutions International BV

Badhuisweg 3

Amsterdam

the Netherlands

e-mail: Fred.Hoeve@shell.com

Abstract: Shell Global Solutions Statistical Consulting and the Materials & Inspection business groups deliver technical services, consultancy and R&D to both Shell and third party customers in the areas of manufacturing process solutions, non-intrusive inspection, maintenance planning and integrity quantification.

Last year a large project was started to upgrade pipework integrity on offshore platforms. To identify pipe sections that need replacement, extensive inspection was carried out, followed by analysis of the data to assess remnant life. Challenges faced during the project were the handling of vast quantities of wall thickness data containing various types of data errors, taking into account different corrosion degradation mechanisms and the sometimes limited inspection coverage typical for pipework inspection.

This paper focuses on the different data analysis methods that have been used during the course of this project, most of them had to be tailored for the purpose and have become part of the Shell Inspection Design Analysis and Plotting (S-IDAP) package. The paper particularly discusses the functionality of an advanced method for trending wall thickness data and various visualisation tools to help understand the distribution of corrosion in the circuits. Further tools were developed to optimise the inspection coverage and to estimate corrosion rates together with statistical confidence measures, providing input into the updating of remnant life parameters of the Risk Based Inspection scheme.

Key Words: pipework, risk based inspection, non destructive testing, non-intrusive inspection, ultrasonics, radiography, corrosion rate trending, inspection database

1 INTRODUCTION

Last year a large project was started concerning the upgrading of pipework integrity on offshore platforms in the North Sea. This project includes an extensive upgrade of the pipework that has been in service for anything up to 25 years. To upgrade the overall integrity, sections of pipework that are damaged as a result of internal or external wall loss have to be replaced. The criterion for replacement is that pipework is required to have a remnant life of at least 5 years, measured against a specified minimum required thickness well above the pressure containing thickness.

The pipework on a production platform is divided in some 15 functional groups and further divided in some 60-120 pipe circuits. Each circuit consists of 50-500 pipe features in the form of bends, straights, reducers, Tee's, etc. Depending on their size such features can be divided in one or more inspection locations. NDT measurements have been recorded in a pipework database since 5 years, with older data summarised as a starting thickness estimate. This has resulted in a volume of up to 10.000 inspection data points per platform.

Coverage of the pipework by inspection is influenced by several factors. Changes in the inspection philosophy over time have resulted in different types of data; older data represent so-called key points (RT-shot; UT-measurements at cardinal points; or scanning of a larger area), whereas more recent inspection procedures require scanning a 200 mm wide band every meter of pipe and full or extensive coverage of bends, Tee's etc. Working practices developed by inspector and corrosion engineer are the basis for selecting a certain proportion of the features within a pipe circuit. For the project a target for the initial level of coverage was set at: 60% of the features and 10% of the welds.

The inspection techniques used in this project are mainly ultrasonic normal beam scanning and gamma-radiography. For welds angle beam ultrasonics and radiography are used, while in a minor number of cases use was made of a Time of Flight Diffraction technique (welds) and MFL/SLOFEC inspection (for rapid scanning of straight sections).

Any discrepancy in the data used for a remnant life calculation will immediately translate in odd, extreme or missing corrosion rates or remnant life estimates, which disturb a clear view on the corrosion behaviour in the circuit, the level of inspection, etc. Therefore the data in the RBI-Pipework database have to be of very high quality. For the remnant life calculation the following data are critical: measured wall thickness, pipe diameter, nominal thickness, minimum required thickness, together with the piping class connecting these three items, and the commissioning date or the replacement date of sections of it. These data have to be connected, via unique numbers, to the features in an isometric drawing, and to the correct piping class and operational parameters as specified in P&ID drawings.

It soon became clear that the data from the routine, day-to-day, inspection program were good enough for a first indication, but insufficient to enable the calculation of remnant life estimates with the required accuracy. This can be explained by the fact that the first aim of day-to-day inspection is to detect damage close to the critical level, whereafter this particular damage is monitored until the replacement level is reached. A risk-based inspection plan, requiring an accurate medium-term remnant life estimate, on the other hand requires accurate wall loss measurements over a larger range to enable accurate estimation of the corrosion rate.

To fulfil both the short-term requirements and the longer-term requirements of the risk-based inspection plan, the following approach was chosen. The available inspection data were both used to find near-critical damage (the usual short term requirement) and to make a first pass assessment of remnant life. With these first estimates dedicated inspection plans could then be designed and carried out to firm up the initial remnant life estimates.

Other challenges to this approach became soon apparent:

- The quantity of data available from an offshore platform is far too large for manual processing
- Existing inspection databases are incapable of providing the required detailed overview of the condition of the pipework
- No reliable method to calculate 'optimal' inspection coverage was available. Particularly challenging in this respect is the fact that corrosion damage in pipework is often very local (bends, welds, low points, etc.)
- Inspection drawings (isometrics), indispensable for planning the execution of inspection, were available, but only in various different formats (e.g. hand-drawings, CAD drawings).
- Longer term historical data series, necessary for estimating corrosion rates, are often curtailed e.g. when pipe sections are replaced or inspection management databases are changed.
- Remnant life calculations are very sensitive to discrepancies in the data, necessitating extensive data quality testing.

It is clear from the above that a better, automated, way to analyse the data was the first requirement. A dedicated computer program was developed and called the Shell Inspection Design Analysis and Plotting (S-IDAP) package. This package can be easily linked both to Excel and to inspection databases. Its use in analysing inspection data will be described in this paper.

2 ANALYSIS OF INSPECTION DATA

After storing the measured wall thickness data in the database the following data processing and analysis steps have to be carried out:

- Estimate the corrosion rate per location, based on the historical wall thickness data.
- Analyse the corrosion rates for each corrosion circuit. The corrosion circuit initially is the smallest unit in the pipework for which similar type and level of corrosion damage is expected.
- Check whether the corrosion circuit indeed is the smallest unit in the pipework. If corrosion damage is more severe, or of a different nature, in particular locations, the corrosion circuit has to be broken down in several parts. This process is called stratification, and the different parts are called (corrosion) strata.
- Decide whether enough measurements have been taken to achieve an adequate confidence level that the corrosion behaviour is represented reliably. If not, make a plan for additional inspections.
- Update the RBI corrosion rate and confidence rating.
- Calculate the remaining lifetime of each piece of pipework (or stratum).

These analysis steps and the use of the S-IDAP package are briefly described below.

2.1 Estimating the corrosion rate per location

Knowledge of the local corrosion rate at each inspection location is essential for the understanding of the corrosion behaviour in a pipe circuit. By looking at actual wall thickness readings over several years, one realises that relative large errors can occur in NDT data, that the corrosion rate is in general not constant over time and that the number of measurements in a series is very limited. These data are therefore not an ideal basis for forecasting using conventional linear regression techniques. Hence we used our statistical trending software tool (TST). This tool, based on Bayesian techniques, deals with anomalies in the data in a natural way, predicts wall thickness and corrosion rate over time and gives an estimate, with confidence bounds, of the remnant life of the pipeline, vessel or other structure under consideration.

The method consists of simultaneously fitting four different models. In the main model (1) the wall thickness and corrosion rate are allowed to vary slowly over time. Furthermore three alternative models (*Figure 1)* are fitted to cope with the following anomalies in the data:

- outliers, a wrong measurement (model 2);
- change in wall thickness, e.g. occurring due to a wrong instrument calibration or a change in paint thickness (model 3);
- change in corrosion rate, a genuine change, e.g. caused by a change in operating conditions, that need to be detected and taken into account as early as possible (model 4).

The posterior probabilities of the four models, given the data at a particular time instant, are calculated and a weighted mixture is used for prediction.

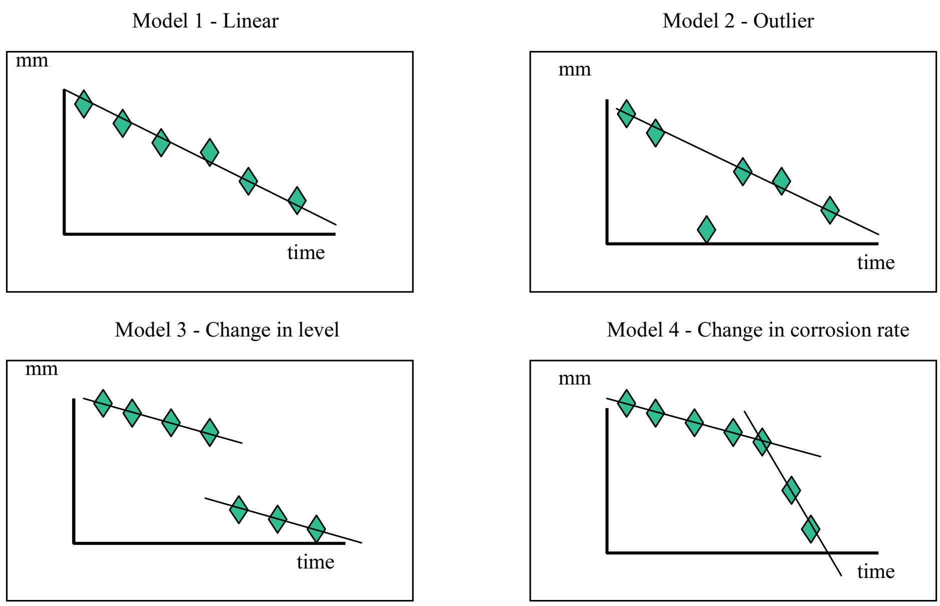


Figure 1, the four models applied in the trending software tool

With the trending software tool large numbers of data are processed automatically and only those locations where an anomaly in the data (model 2, 3 or 4 prevails) has been detected are selected for manual review. A replacement of a pipe spool, which is not documented in the inspection database, will show up as a discontinuity (change in level). An example of the trending graph is given in figure 2. In this example the dark blue line at 11.7 mm represents the minimum required wall thickness and the blue points are wall thickness measurements. The predictions are shown as purple points, the accompanying error band can be interpreted as a prediction interval for a single future measurement. It is based on the updated confidence in the model and the supplied accuracy of the UT wall thickness measurements. The red rectangle at February 2003 indicates that this measurement is detected as an anomaly. This has been confirmed by the inspector. The estimated corrosion rate of 0.14 mm/year is used to calculate the remnant life for this location (i.e. feature A12).

The TST package has been developed under the framework of Shell Risk Based Inspection (S-RBI) and is routinely used at a number of Shell refineries and upstream production sites.

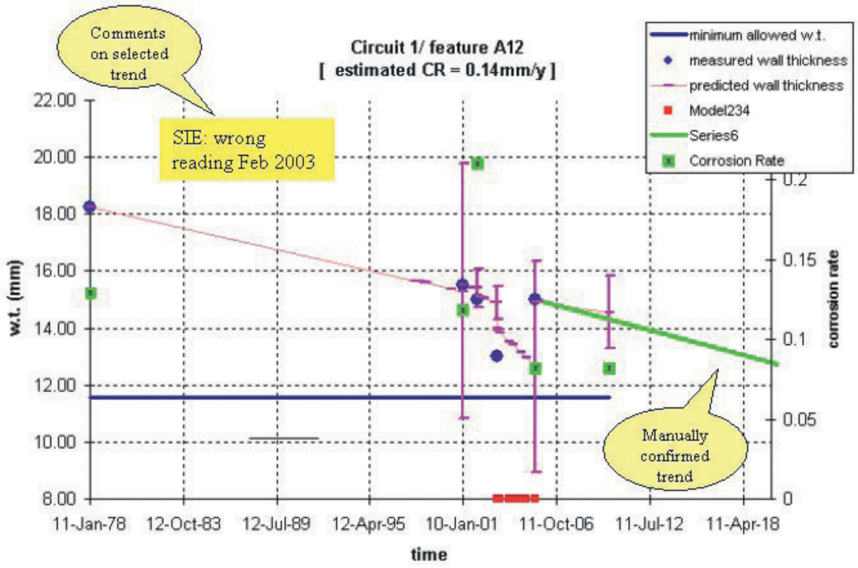


Figure 2 Graphical output of the TST package. A wrong measurement has been detected Feb 2003.

2.2 Analysing the corrosion rates for each corrosion circuit

After trending the wall thickness data at each location, we continue the analysis by displaying the corrosion rates in different plots per corrosion circuit, which is initially the smallest unit in the pipework for which a similar type and level of corrosion is expected. With the help of the plots the corrosion engineer can judge whether an update of the circuit corrosion rate (i.e. the RBI corrosion rate) should be made, to better represent the actual corrosion rate of the circuit. For some circuits a further breakdown of a corrosion circuit is necessary, this process is called stratification. In total about ten different plots can be produced in S-IDAP. Examples given in the following figures are the histogram plot, the box-plot and the overview plot.

In Figure 3 the histograms for two circuits, 14 and 12 are given. The first histogram shows a low corrosion rate for the whole circuit, while the second shows that higher corrosion rates are also present. Circuit 12, for instance, needs to be further analysed to decide whether further stratification within this circuit is necessary. In Figure 4 a categorised box plot is shown. From this plot it is clear that corrosion rates in bends on average are considerably higher than those in the other features, requiring a split up of this circuit into 2 corrosion strata.

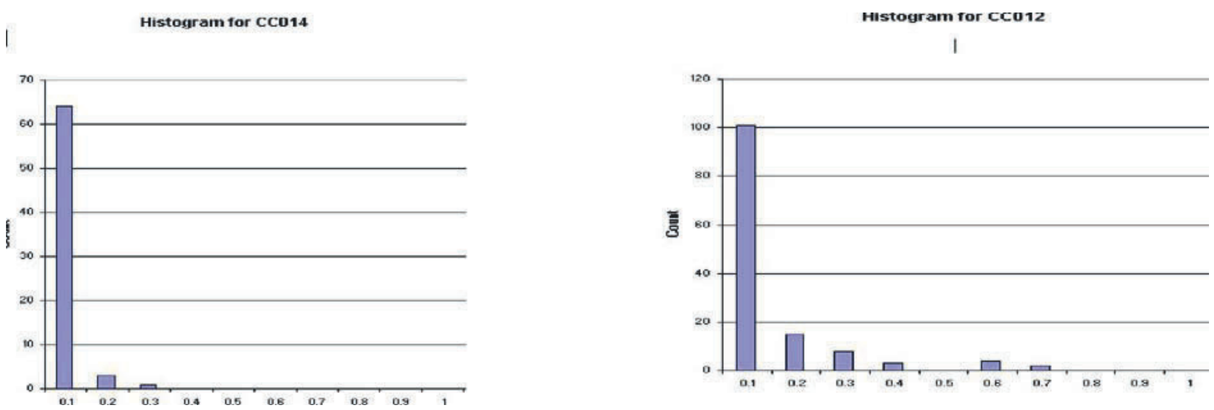


Figure 3 Histograms of corrosion rates from 2 circuits

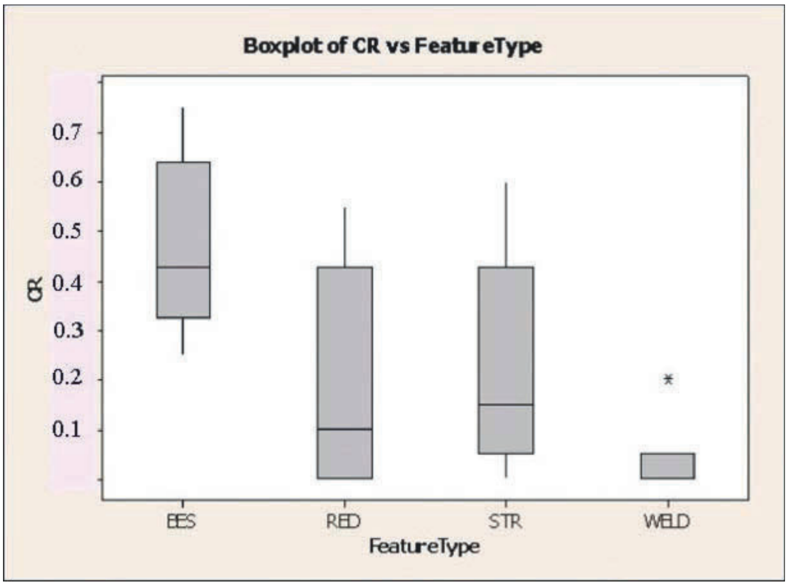


Figure 4 Categorised (per feature type bend, reducer, straight, weld) box plot of corrosion rates

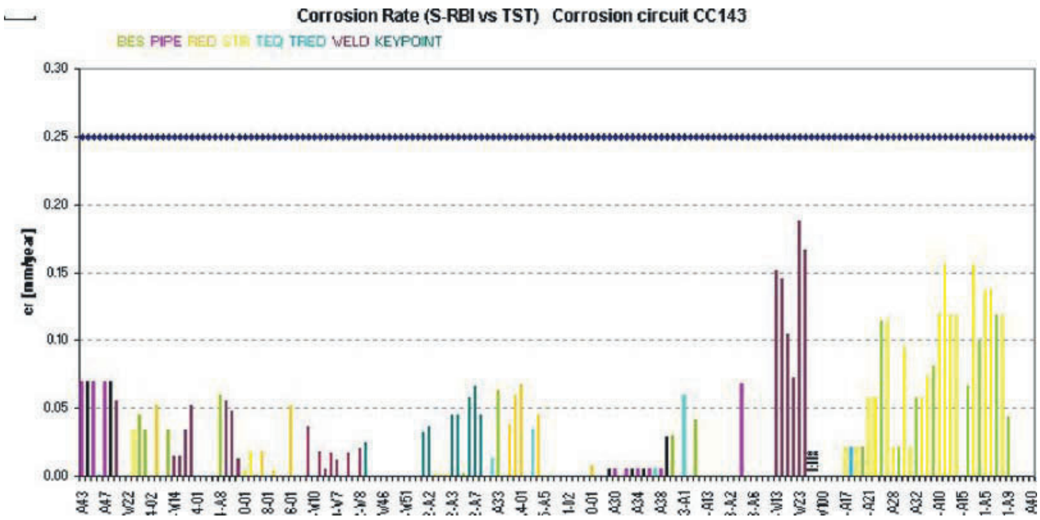


Figure 5 Circuit overview plot. Corrosion rates of adjacent features are shown, colour-coded per feature type

Figure 5 shows a circuit overview plot. The corrosion rates of all features in the circuit are shown, colour-coded per feature type. The dotted line is the RBI-selected corrosion rate for the circuit. From this plot it can be easily seen whether certain features show different corrosion behaviour or whether corrosion is more severe in certain parts of the circuit. From this overview of measured local corrosion rates the RBI-selected corrosion rate can easily be updated. The overview plots are automatically produced from data down-loaded from the pipework database into spreadsheets and calculations are carried out using the S-IDAP library and applying the Shell-Non-Intrusive Inspection methodology.

2.3 Colouring hand-sketched or autocad isometrics

Further understanding of the behaviour of the corrosion mechanism(s) can be obtained by marking up isometrics with the corrosion rate at each measured location (Figure 6). Such marked-up drawings provide an overview at a glance of the locations where high corrosion rates occur. This often supports the explanation for such extreme corrosion rates, for instance because a feature is a dead leg or a high-flow area. Data points that possibly contain an error are specially marked so as to avoid false interpretations because of data errors. The drawings are also used to display other information, such as remnant life, date of inspection, or wall loss. It can also give a good indication for the necessity of further stratification.

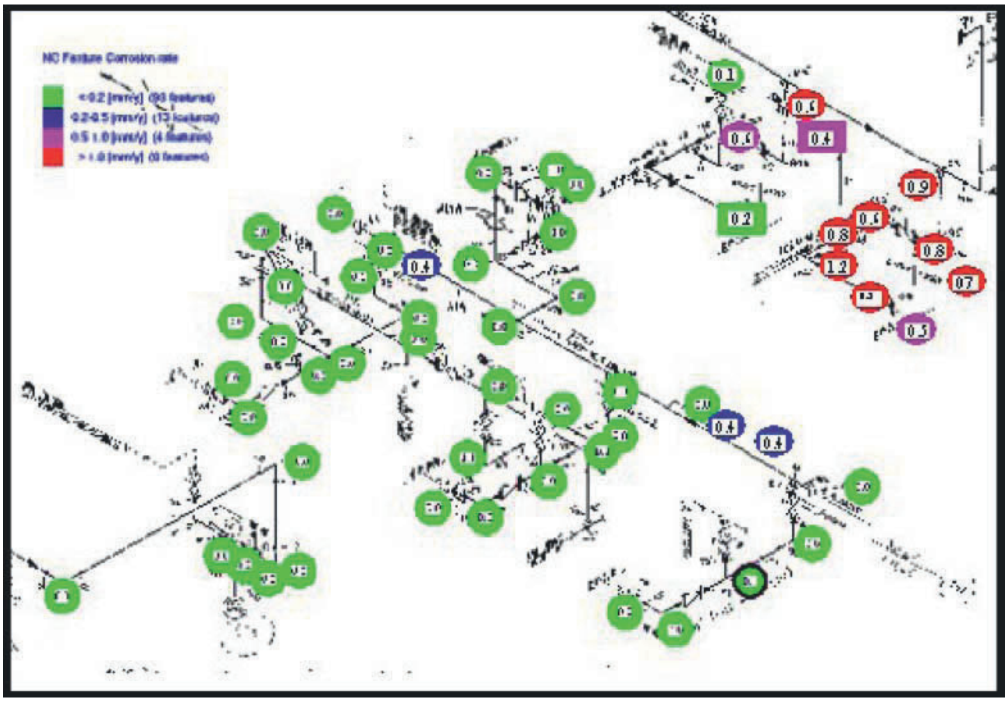


Figure 6. Isometric drawing of a circuit, marked up electronically with the inspection points, and displaying information about the corrosion rate, (by colour and number). The symbols identify locations at a dead leg (rectangle) and data anomalies as indicated by the trending package (black border).

2.4 Stratification of corrosion circuits

If preferred locations for corrosion occur in a circuit, this may result in an adverse effect on planning and effectiveness of the inspection. Firstly, when selecting the highest corrosion rate anywhere in the circuit for calculating the inspection interval, this will generate a large penalty in the form of unnecessary, premature inspection in a large part of the circuit. Secondly inspectors, with support of the corrosion engineer, often build up knowledge over time of the preferred locations for corrosion, so that inspection effort can be directed towards these locations. However, if the administration of this knowledge is not supported by proper functionality of the inspection database it may become subject to an individual's judgment, prone to human errors, and it may easily be lost over time due to staff changes. To make this process more objective and open for verification use can be made of sampling rules; as these rules are strongly driven by the estimate of the largest corrosion rate, they are very sensitive for proper "stratification" of a circuit, i.e. the subdivision in hot spots and the rest.

Stratification can be visualized using isometric drawings marked up with corrosion rates (as in Figure 6) and further analysed by splitting the circuit in corrosion strata and studying the corrosion rate histograms (Figure 2) or box plots (Figure 3).

2.5 Analysing inspection coverage requirements

In practice it is often impossible and normally undesirable to inspect 100% of a structure such as a piping system. Statistical sampling techniques allow the necessary inspection coverage to be defined and the results of a limited inspection to be extrapolated to the status of the total system. S-IDAP has functions that calculate either the sample size or the coverage that has to be inspected to achieve a certain confidence level. The required input consists of the RBI parameters (next inspection interval, confidence rating for the corrosion rate, criticality) and the expected corrosion morphology, general (uniform) corrosion or localised (pitting, grooving) corrosion.

This set of functions has also been developed under the framework of Shell Risk Based Inspection (S-RBI) and has been used at a number of Shell refineries and upstream production sites.

By its nature, pipework introduces complexity in the assessment of coverage: each circuit is made up of parts with a different corrosion allowance, whereas corrosion happens locally within the circuit but may be expected to occur anywhere. This requires that the components in pipework need to be organized along two dimensions: in groups of equal corrosion allowance, and in circuits subject to similar degradation.

2.6 Calculating the remnant life

After the calculation of the corrosion rate and the minimum allowable or required wall thickness the calculation of the remnant life is straight forward. By displaying the remnant life in the hand-sketch or autocad isometric one gets a good overview of the problem areas in the pipework.

3 SHELL-INSPECTION, DESIGN, ANALYSIS AND PLOTTING (S-IDAP) LIBRARY

The functions described in this paper are all present in the S-IDAP library. Furthermore, the package provides routines that can give an overview of many plots on one screen, see e.g. Figure 7 below. In this figure a histogram is displayed in the upper left part, an overview plot in the upper right part, a trending plot in the lower left part and the isometric in the lower right part. After an update by the user in the trending plot the other plots will automatically be updated. By selecting another feature in the isometric the corresponding trending plot will be selected. Further, a polygon can be drawn around a number of features to define a new corrosion stratum.

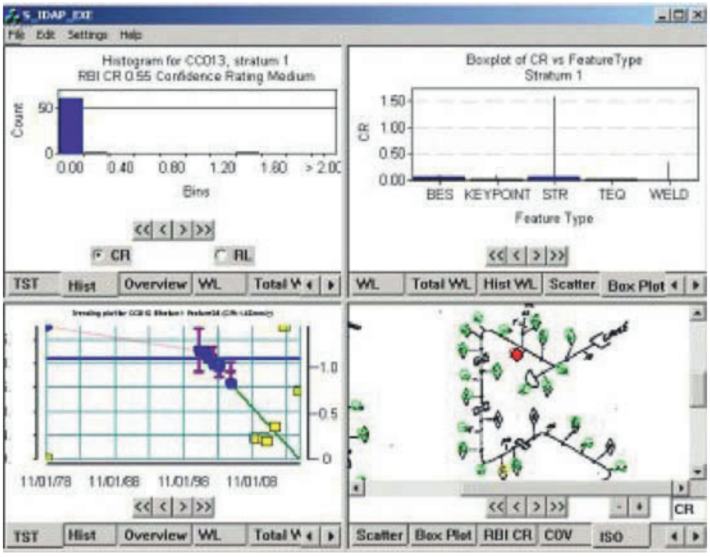


Figure 7. Overview of different plots in one go. The high corrosion rate showed in the histogram is the red dot in the isometric. The corresponding trend plot confirms that corrosion rate is really high !

4 CONCLUSIONS

During the course of performing a large-scale project to upgrade pipework on offshore production facilities a number of problem areas and learning points arose. The main conclusions can be summarised as follows:

- When the planning of inspection and repair activities assume long remnant life, a strong focus is required on obtaining accurate corrosion rate estimates, not just on finding near-critical defects.
- Pipework inspection databases have to be modified to support the analysis of corrosion rates more appropriately. For better inspection planning, combining sections with similar corrosion behaviour and pipe-sections with the same tolerance should be possible.
- New inspection data should be assessed and validated immediately following the inspection, enabling a better judgement of its consistency, enabling better planning of further inspections.
- Inspection staff should be trained to understand the methodologies of assessment and to operate the TST tool.

These findings, together with the fact that large amounts of data had to be analysed, lead to the development of the Shell Inspection Design Analysis and Plotting (S-IDAP) package. This package is composed of a combination of existing and newly-developed tools, brought together in a package that can be easily linked to Excel and to inspection databases. The package already showed its value during the course of this project. Its main functionality can be summarised as:

- A tool for trending wall thickness data (TST), which deals with data anomalies and changes in corrosion rates. With TST large amounts of data can be handled automatically, so that manual attention is only required for special cases.

- A comprehensive set of plotting routines, assisting the user with data validation and interpretation. Some of the plots are also valuable in helping the user with the difficult, but important, process of stratification of pipe-sections.
- A tool to aid the user in designing 'optimal' inspections within a stratum, with respect to the required statistical confidence in remnant life estimates.

Acknowledgments

The authors would like to thank the staff in the Mature Asset team and the Upgrade Project within Shell Exploration and Production Europe, for their cooperation and their support to the implementation of the inspection strategies.

SYSTEM ARCHITECTURE AND DATA DESIGN FOR RFID-BASED ITEM LEVEL TRACK AND TRACE

O.P. Gan, Y.Z. ZHAO, M. LUO and J.B. ZHANG, J. H. Zhou [a]

[a] *Singapore Institute of Manufacturing Technology, 71 Nanyang Drive, Singapore 638075. Tel: +65 67938993. Fax: +65 67916377. Email: opgan@SIMTech.a-star.edu.sg.*

Abstract: This paper presents a system and methods for tracking and tracing inventory items in retail stores and manufacturing warehouses. The motivation is to provide businesses in retail and manufacturing industries with an effective technology using Radio Frequency Identification (RFID) to practice lean inventory management. The current focus in major RFID trials around the world is steered towards at pallet and case levels. Using RFID to track and trace inventory at item level is still at its infancy stage. The challenges of item level RFID implementation are more complex than at pallet level particularly in the area of data management and device management. This paper presents an architecture design of a networked RFID tracking and tracing system, and also proposes a data schema design for managing track and trace data.

Key Words: Radio Frequency Identification, Middleware, Track and Trace, Item-level Inventory Management, Data Management.

1 INTRODUCTION

Lean inventory management is important to many industries including manufacturing, retail, logistics and healthcare. This is more so in today economy in which companies compete to offer products and services at the best possible values at the lowest possible cost, and to be quick in adapting and responding to the dynamic change in market demands. To be lean in inventory management, there is a need to track every item in the inventory. This requires the capability of telling the current location of every item, how long in time each item has stayed in the store and the history of location and time for every location each item has moved to. Such requirements are difficult to achieve using the mature bar-code technology with manual stock taking process. The current approach is labor intensive, time consuming and highly prone to error. In many retail stores, inventory items are managed at stock keeping unit (SKU) level. The lack of item serialization presents another challenge to track the age of individual items.

RFID is widely accepted by industries to be the emerging technology for product identification [1]. In fact, multiple RFID trials for warehouse applications are already underway in various parts of the world. The well-known case is Wal-Mart RFID mandate of 2005. Other similar efforts are also reported in Carrefour in France, Marks and Spencer in United Kingdom, Metro AG in Germany. Most of these efforts focus at pallet level in warehouse applications. A study reported in [2] discussed about the use of RFID in air cargo tracking in which air cargo pallets were tagged and RFID readers were installed at material handling vehicles in the airfreight terminal for tracking the pallets within the warehouse.

The challenge in item level track and trace is to pinpoint the current location as well as the historical movement of every item in an environment. In many retail stores, this information is useful especially in stock taking and sales operations where finding stocks indicated as available in system is often a daunting task. A frequent reported problem in retail stores is inefficiency and inaccuracy in manual stock taking process [3]. Usually manual stock taking is carried out periodically from weekly to monthly or even half yearly. In every stock taking exercise, the aim is to close the gap between physical count and system count. The gap will tend to grow as the interval of stock taking is further apart. This has an effect on purchasing function which is often based on system count and not physical count. An undesired effect is to order more than necessary or order less than required. The problem with the former is excess stock and depreciation, and with the latter is loss of sales due to stock shortage. By tracking the location and time of every inventory item, the problem of instant stock counting is also solved. In fact, this paper shows that such a track and trace system can provide a more detailed stock count with location information.

The problem of aging inventory is another issue that can be solved by a track and trace system. An inventory age refers to the duration the item stays in a store from the time it is delivered from supplier. The current practice in some retail stores is to estimate the age of inventory based on delivery order records. Such estimation is good for indicating general inventory age. However it is not good enough to tell which item among a group is older. Tracking inventory age is important especially for perishable goods as well as goods with high depreciation. In retail industry, price competition is normal among retailers. Very often, items that are older are preferred to be sold at discounted price than new arrivals. This presents a requirement to identify and locate aging inventory items. With the proposed track and trace system, this requirement can also be met.

Pilferage detection is another important task in retail industry. A common approach is using the 1-bit electronic article surveillance (EAS) technology [4]. An EAS system comprises a tag and a gate reader. When a tag is not disabled, it will trigger a security alarm system when it passes through the gate. EAS is commonly used together with barcode technology. It is not uncommon to find two different labels on an item: one for EAS and another for barcode. The former provides security function while the latter provide identification. RFID can be used to provide the two functions in one system.

Read reliability and data scalability are two problems that make item level tracking and tracing more difficult to handle than pallet level. Read reliability refers to the ability of an RFID system to successfully interrogate every item without a miss. Perfect read reliability requires every item to be identified without any miss-read at all times. However, this is difficult to achieve in practice due to radio frequency blind spots and interference from external sources such as human crowd. One way to reduce miss-read is by deploying multiple layers of RFID interrogation zones with complementary reading functionality. In this area, the idea of networked RFID is discussed in this paper.

In item level tracking, each item has a RFID tag containing a unique identification code. As the size of inventory increases, the volume of data generated in an RFID system is expected to grow exponentially since a single tag can be read multiple times at various locations. Data scalability refers to the problem of managing large volume of data in the more efficient manner. Poor data management is undesirable since it can result in memory overflow, loss of data and slow data processing. One way to streamline data management is by designing a compact data model. In this area, a novel data schema is proposed in this paper.

The contributions of this paper include an architecture design of a networked RFID track and trace system and a data schema design for tracking inventory items. The rest of this paper is organized as follows: Section II describes the networked RFID track and trace system. Section III presents an approach for data schema design. Finally, Section IV concludes.

2 NETWORKED RFID TRACK AND TRACE SYSTEM

Every item to be tracked and traced requires a unique identity code. This is commonly known as product serialization. An emerging standard for product serialization is the Electronic Product Code (EPC) [5]. The standard specifies the use of 96 bits comprising manufacturer code, product type and serial number. A RFID tag stores the identity code. In the EPC standard, there are several classes of RFID tags. Each class differs in features and performance. Class 1 Generation 1 and Class 1 Generation 2 tags are two most widely mentioned tags now. Class 1 Generation 2 offers better features such as longer read range, better anti-collision and faster speed than Gen 1. Cost of tag is still fairly high ranging from US 30 cents. As such, the cost of tagging every item in an inventory can contribute a large running cost of the system.

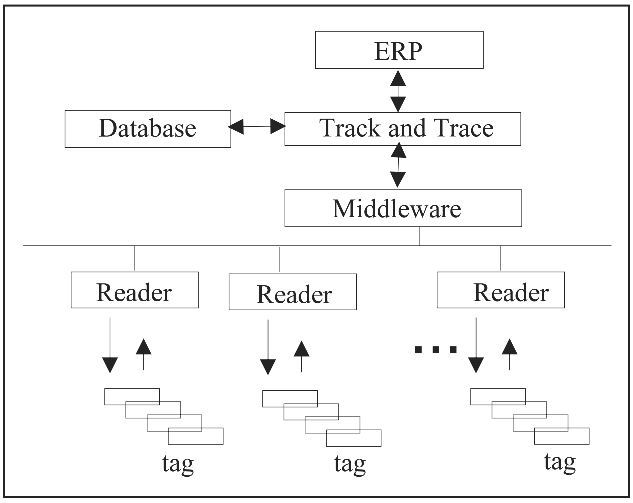


Figure 1: Track and Trace System Architecture

With every item given a RFID tag, a system is required to read the RFID tags in order to identify tagged objects and track its movement as they are moved from one location to another in an environment. The system should comprise a network of RFID readers deployed across the environment. Fig. 1 shows the networked RFID system architecture which comprises RFID tags, readers and middleware, database and track and trace application. RFID readers can be flexibly deployed in various strategic locations in the environment including exit and entrance points, checkout points, path, shelves and open areas. Three levels of granularity of RFID coverage can be considered: 1) Exit/Entrance coverage 2) Boundary coverage and 3) Point coverage. The choice of granularity depends on specific application needs. Exit/Entrance coverage is the basic option. It covers the exit and entrance points of an environment as well as check out stations. In a retail store, exit and entrance points can be found in the showroom as well as in the receiving store. Deploying RFID readers at exit and entrance points is sufficient to achieve stock counting, aging inventory monitoring, pilferage detection as well as limited location tracking and tracing.

Where it is useful to know the distribution and tracing of item movement across departments, boundary coverage should be considered. This concept involves the deployment of RFID readers along department boundary. It requires more reader investment than the basic option. The gain is additional information including stock counting by locations and tracking of items at department level. Point coverage offers the highest granularity among the three. This refers to the large scale deployment of RFID readers at strategic points in the environment where items are stacked or shelved. In retail stores, it is quite common to find stacks of products on the floor. With point coverage, a RFID reader is deployed above each stack as shown in Fig. 2 so that item taken out or placed on the stack can be tracked. This increases the investment on RFID readers.

The gain is pinpoint tracking of each item with higher resolution. Aging items within a stack can be identified and locating items is also much easier.

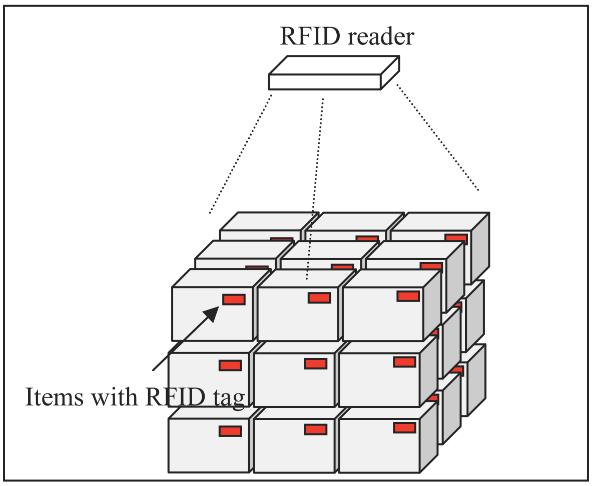


Figure 2: Stack Item RFID Readability

Read reliability is not 100% particularly for stack items. Referring to Fig. 2, the top row of stack items usually has no problem with readings. However, items in the middle and bottom rows may not be completely read at all times. These items can be read when they are removed from the stack.

Above the RFID readers and tags in the system of Fig. 1, is the RFID middleware layer and Track and Trace application. RFID middleware manages and controls RFID readers. Passive tags do not have battery. In order to power up its integrated circuit, a passive tag obtains energy from RF wave produced by a RFID reader. After power up, a tag proceeds to transmit data to reader. Data received at each reader is first stored in the reader memory and subsequently transferred to the middleware when a reader is polled. Data loss due to reader memory overflow is possible if a reader is not serviced in time by the middleware. This is possible to occur in two ways: a) new data is rejected b) old data is deleted to make room for new data. To avoid reader memory overflow, it becomes critical that the Middleware controls the polling of readers in an optimum way such that each reader will not be starved of data service opportunity. Some RFID readers in the ultra-high frequency band including that of Alien and Symbol are offering memory capacity about 1000 tag data. The network that links every reader to the middleware can be seen as a shared media. At any one time, only one reader is allowed to transmit data to the middleware in order to avoid data traffic collision. The reader allocation function in the middleware runs an assignment algorithm to decide the next reader to allow transmission. Reader assignment problem in RFID middleware is similar to the classic polling problem. Static allocation algorithm based on fixed time slice sharing is widely reported in literature [6]. The static algorithm is useful for the case with similar data traffic load condition in every reader. However, it is possible for cases with wide difference in data arrival rate at each reader to occur. For this situation, the static algorithm of reader assignment is not efficient. What is needed is an adaptive algorithm that assigns the right to transmit to the reader according to the data traffic load of each reader. The reader with the heaviest traffic loads should be served first to avoid memory overflow.

Large volume of data is collected by the RFID middleware from a network of readers and processed in the track and trace layer. According to Fig. 1, the track and trace layer lies above the middleware layer. Application specific customization is allowed in the track and trace layer and does not need to be in the middleware layer. In addition, the track and trace layer provides the graphics user interface for stock counting information and aging inventory information. Tracking and tracing information is presented in a graphical map window of the environment as shown in Fig. 3. Flashing circle represents the current location of an item. Filled circle represents historical locations visited by an item. Lines with arrow represent the path taken by an item. As an item is moved from one location to another, the graphical map will be updated with new circles. A listing of the historical locations and timestamps is also provided in a separate window. Another useful feature is the use of different color codes to display recent tag readings and old readings. Using a user defined time threshold e.g. 5 seconds, recent tag readings are displayed in green color and the older readings are displayed in red color.

By maintaining a set of track records and trace records for every item, the track and trace layer continuously updates the location of every item and also shows the route of movement for individual items. Additionally, the track and trace layer performs stock counting and monitoring functions. Items recorded in the database are retrieved for stock counting information by adding all items according to SKU and locations. The age of an inventory item starts from the time the item arrives at the store. The track and trace layer maintains a set of timestamps for every item. One of the timestamp is initialized with a time corresponding to the first moment the item is received and read in the receiving store. The age of an inventory item is computed as the different between current time and the timestamp. Monitoring stock levels is another key function of the track and trace layer. This refers to the task of highlighting specific stocks in which the current quantities are below the minimum quota required as well as stocks that are more than the maximum quota allowed. This information is useful to support purchase decision in enterprise purchasing department. In fact, the track and trace layer could feed stock counting and monitoring information directly to the enterprise resource planning (ERP) system (also shown in Fig. 1) to improve the efficiency of purchasing.

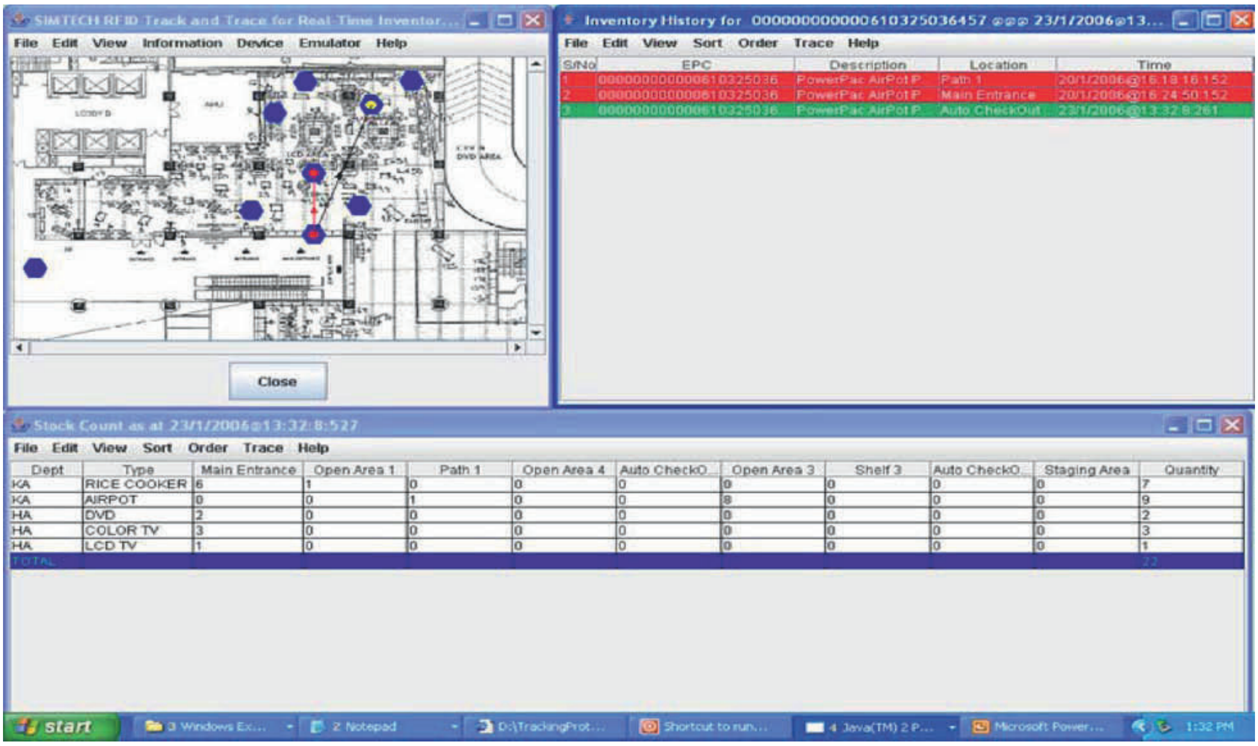


Figure 3: Stock counting, item tracking and tracing windows

Detecting and alerting pilferage is another function that the track and trace layer should provide. By maintaining a checkout status of every item, the system triggers an event to activate the enterprise security alarm system when an item without proper checkout code is identified at an exit location. In addition, the system provides a full detail of every occurrence of pilferage event including location, time and item identity. To provide the inventory related information stated above, an efficient data model customized for track and trace is needed. The next section gives a detail account of the proposed idea.

3 DATA MODELING AND UPDATING

Every item has a unique identity code encoded in the RFID tag. As item is moved from one interrogation zone to another within the network, the identity code from a tag is captured by RFID readers at both source and destination locations. A reader normally timestamps every reception of identity code based on its internal clock. The identity code and reader clock timestamp data are stored in the reader internal memory and subsequently sent to the middleware when the reader is polled. The middleware should also timestamp data from readers based on the middleware clock. In summary, item identity code, reader timestamp and middleware timestamp are data processed by the track and trace layer for every tag reading before storing in an item record in database.

Fig. 4 shows a schematic view of an item record in the database. Each item record includes an identity code field, a state field and a track record pointer and a trace record pointer. The state field includes several values including Available, Checkout, Pilferage, Check-in, Return and Servicing. The track and trace layer updates the state of each item according to the state transition diagram in Fig 4. When suppliers deliver items to a retail store, the state of each item is assigned with the value of Available. The state is marked Checkout in the state field when the items are sold. If items are detected at exit locations but not checkout, the item is marked pilferage in the state field. When customers return items, the items are marked Return. When items are sent for servicing, the state is changed to Servicing. After servicing, the items are ready for sales and marked Available.

Fig. 4 also shows a schematic view of the track record and trace record. Both types of record contain a plurality of location records. A location record contains several fields including location code, first timestamp and last timestamp, reader timestamp and middleware timestamp. The location code identifies every interrogation zone in the network. The first timestamp corresponds to the first RFID reader timestamp captured at the location for the specific item. This timestamp is not changed after it is created. The last timestamp is the most recent RFID reader timestamp captured at the location for the specific item. This timestamp is updated whenever there is a new timestamp from the RFID reader. The time clock in every RFID reader in the network should be synchronized with the RFID middleware.

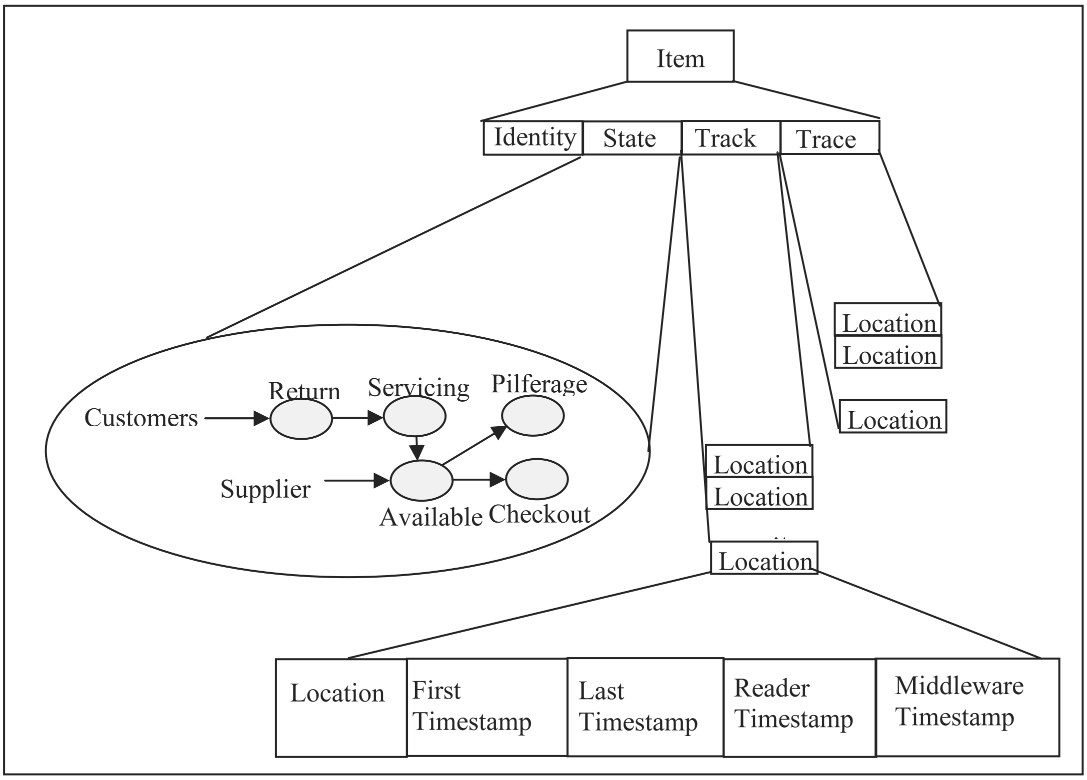


Figure 4: A schematic view of item record

The track record contains a unique set of location records such that one cannot find two or more location records with the same location code. When an item is moved to a new location for the first time, the track and trace layer adds a new location record to the track record. In subsequent visits to the same location, the track and trace server does not add new location records in the track record. Instead, it updates the timestamp fields of the specific location record with the most recent timestamps. The location records in the track record are also never deleted unless the item record is removed. With this arrangement, the track record of an item shows the most recent timestamps for all the locations visited by the item. The most recent location visited by an item corresponds to the location record with the most recent timestamp in the track record.

The trace record contains a set of unique location records such that it is possible to find multiple location records with the same location code but not the same last timestamp. The trace record is for the purpose of tracing the history locations of every item in the inventory. The track and trace layer stores every location visited by each item up to the present location, and adds a new location record in the trace record for every new location visited if the new location is different from the last recorded location in the trace record. It also updates the timestamps of the most recent location record if the item is identified to be at the same location as the most recent location record. This can happen when items are stationary at an interrogation zone and continuously being identified by the RFID reader. With exception to the most recent location record in the trace record, all other location records cannot be modified once added to the trace record. However, the old location records can be deleted when the number of location records grows beyond a user-defined limit. The location records in the trace record correspond to the location trace of the item.

It is possible for no read to occur due to radio blind spot and other environment interferences such as human crowd. In this case, items affected cannot be identified in the interrogation zones. This is particularly so for a stack of items in which items in the topmost layer can be identified but not those in the middle and bottom layers. Usually items at the middle and bottom layers cannot be removed unless those at the top layer are removed first. As such, those items in the hidden layers are likely to be identified. The track and trace system is designed to update the location records only if RFID readers can identify items in an interrogation zone. If no read occurs, the location records are not updated. This does not cause error in tracking if the item remains in its present location. An error occurs if the item is moved to another location without being identified. The system reduces this error by employing a network of interrogation zones such that items that fail to be identified at one interrogation zone is likely to be identified at other zones. As the granularity of the interrogation zone network grows, accuracy of tracking also increases.

4 CONCLUSIONS

Growth of interests in RFID application across many industries is ever increasing. Pallet level tracking is becoming mature soon. The next focus is steering towards item level tracking and tracing. This is especially useful in retail industry where inventory management is still a highly manual process. To close this gap, a system and data design for a RFID-based track and trace system are presented in this paper. The main contributions of this work includes a design of the track and trace layer above the middleware layer in the proposed architecture, as well as a design of an item record with track record pointer and trace record pointer. Each pointer is linked to a plurality of location records comprising of timestamps and location.

As part of future work, an adaptive reader assignment algorithm based on estimation of individual reader data traffic condition will be investigated. This is needed to ensure that all readers are serviced without loss of data. In addition, an information reliability model will be derived to study the accuracy of tracking information in view of uncertainty in RFID readability particularly in areas where RF blind spots are prevalent. Also, optimization of the RFID reader location with mobile readers will be explored to improve the track accuracy.

5 REFERENCES

[1] Ryo Imura (2005) Driving Ubiquitous Network-How can RFID Solutions meet the customers' expectation. In 10th International Conference on Emerging Technologies and Factory Automation, Italy.

[2] O.P. Gan, T.J. Ng (2005) Radio Frequency Identification Middleware with Fault Detection. In IEEE Conf. on Service Operations and Logistics, and Informatics, Beijing, China, pp. 889- 894.

[3] Moran, Humberto J., Sana Ayub, Duncan McFarlane (2004) Auto-ID Use Case: Improving Inventory Visibility in Retail Company – Impact on Existing Procedures and Information Systems. MIT Auto-ID Center White Paper.

[4] K. Finkenzeller (1999) RFID Handbook. John Wiley & Sons.

[5] EPCglobal Inc™ (2004) The EPCglobal Network™: Overview of Design, Benefits, & Security. EPC Global White Paper.

[6] Hideaki Takagi (1991) Application of polling models to computer networks. Computer Networks and ISDN Systems 22, pp. 193-211.

Acknowledgments

This work was supported by the Agency for Science, Technology and Research (A*Star) under the RFID Proof-of-Concept program.

GENETIC ALGORITHMS FOR FEATURE SUBSET SELECTION IN EQUIPMENT FAULT DIAGNOSIS

Junhong Zhou, Xiang Li, Oon Peen Gan [a] and Shuguo Han, Wee Keong Ng [b]

a Singapore Institute of Manufacturing Technology, 71 Nayang Drive, Singapore 638075. Tel: +65 67938289. Fax: +65 67916377. Email: jzhou@SIMTech.a-star.edu.sg.

b School of Computer Engineering, Nanyang Technical University, Nanyang Avenue, Singapore 639798.

Abstract: This paper presents a method to determine optimum feature subset selection with a modified wrapper-based multi-criteria approach using genetic algorithms. We present details of the algorithm, design and implementation of feature subset selection using genetic algorithms. The best compound features found by genetic algorithms are verified by multiple regression models and are used to construct fault prediction models. A case study of machinery tool wear-out prediction is presented. The fairly good agreement between the prediction result and real tool wear-out data demonstrates the viability of the feature subset selection method for diagnosis applications.

Key Words: Genetic Algorithms, Feature subset selection, Fault Diagnosis

1 INTRODUCTION

In an era of intensive competition, where asset usage and plant operating efficiency must be maximized, unexpected downtime due to machinery failure has become more costly than before. Therefore predictive maintenance has been actively pursued in the manufacturing industry in recent years, where equipment outages are predicted and maintenance is carried out only when necessary. To ensure successful condition based maintenance, it is necessary to detect, identify and classify different kinds of failure modes in the manufacturing process. Sensing technology is necessary in implementing the detecting system. There are two major approaches using sensing technology: the direct method, which measures and evaluates the fault from a single direct sensor, and the indirect method that adopts different kinds of sensors at different positions to sense different failure modes. Features are then extracted and selected to analyse the signals from all these sensors to assess the condition of the system. In many applications, direct method is not possible for on-line realization. Indirect method may work as an on-line systematic technique as it measures the operating parameters during the manufacturing process. In the indirect method, amount of sensors could be large and determining failure could mean taking minutes or hours to analyse data from all the sensors and make a reasonably accurate deduction.

In many complicated applications, it is not unusual to find problems involving hundreds of features. However, it has been observed that, beyond a certain point, the inclusion of additional features leads to a worse performance. Moreover, the choice of features affects several aspects of the recognition process such as accuracy, learning time and necessary number of samples. Most importantly, this leads to an increase in time and computational space complexity of the recognition process. The main goal of feature subset selection is to reduce the number of features used in classification without compromising on accuracy.

Feature subset selection algorithms can be classified into two categories based on whether or not feature subset selection is performed independently of the learning algorithm used to construct the verifier. The filter approach to feature subset selection is done independently of the learning algorithm. Another technique, known as wrapper feature subset selection [1], uses the method of classification itself to measure the importance of a feature or feature set. The first one is computationally more efficient but its major drawback is that an optimal selection of features may not be independent of the inductive. On the other hand, the wrapper approach is more computationally expensive and tends to provide better results than the simpler filter methods [2].

Genetic Algorithms (GAs) [3] can be used for feature subset selection problems since they are generally effective for rapid global search of large, non-linear and poorly understood spaces [4, 5]. The power of GAs comes from the fact that the technique is robust and deals successfully with a wide range of problem areas, including those that cannot be solved by other methods. Over the recent years, many research works have been devoted to the feature subset selection using genetic algorithm for reducing feature numbers without accuracy reduction [6] [7] [8]. These works only study the feature subset selection without further study about their influence on the faults. Little work has been reported on the feature subset selection in the manufacturing equipment fault diagnosis.

The objective of this study focuses on feature subset selection for fault diagnosis through a modified wrapper-based multi-criteria approach using GAs. The feature subset selection results are verified using multiple regression models. The selected feature subset is used to build diagnosis models for fault diagnosis. Implementation issues of GAs on feature subset selection in equipment fault diagnosis are discussed in detail. The feature subset selection and diagnosis system has been tested experimentally with the prediction of tool wear-out on a high speed milling machine.

2 GENETIC ALGORITHMS FOR FEATURE SUBSET SELECTION

Feature subset selection refers to the selection of an optimum subset of features that are most responsible for a given outcome. Feature subset selection problems are found in all machine learning tasks (supervised or unsupervised), classification, regression and time series prediction. The objectives of feature subset selection are to improve the prediction accuracy and to provide a better understanding of the underlying concept that generated the data. GAs is one of the methods that used for feature subset selection [6] [7] [8].

2.1 Genetic algorithms

GAs are inspired by the Darwin theory of principles nature selection and survival of the fittest. The fundamental mechanisms derive the evolutionary process include selection, crossover and mutation within chromosomes. Selection occurs on the current population by choosing the highly fit individuals to reproduce. The selected individuals reproduce new individuals as offspring by crossover with other individuals in the population. Mutation may happen in the reproduction. In this way, over many generations, good characteristics spread in the population, mixing and exchanging with other good characteristics as they go. GAs are computer programs that create an environment where populations of data compete and only the fittest survive, sort of evolution using a computer. GA is well suited for feature subset selection when the number of feature set is large. Though robust, GA takes a high computation time [9]. If GAs is designed well, the population converges to an optimal solution for the problem in a reasonable time.

Generally, there are five steps to implement a GAs:

(1) Randomly generate an initial population.

$X(0) := (x_1, x_2, \ldots x_n)$ of chromosomes

In our case, $x_1, x_2, \ldots x_n$ represents the initial feature sets.

(2) Establish a method to evaluate the fitness $F(x_i)$ of each chromosome x_i in the current population $X(t)$.

(3) Create a new population by repeating the following steps until the new population is complete

i. Selection - Select from the population according to some fitness scheme.

ii. Crossover- New offspring formed by a crossover with the parents

iii. Mutation - With a mutation probability, mutate new offspring at each locus (position in chromosome).

(4) Retain the desired number of fittest individuals in order to maintain the size of the population.

(5) $t := t+1$, if not termination.

The major design components of the GAs include the initialisation process, the design of the evolutionary functions and an objective fitness function. Our GAs design and implementation refer to some of the work in [10].

2.2 Chromosomes creation

One form of encoding is the binary string. The chromosome may be encoded as shown in the Table 1 below:

Table 1

Chromosomes with binary encoding

Chromosome A	1011001010110010
Chromosome B	1111110000011001

Each bit in the string represents one feature. If the bit is 0, it means the feature not selected; otherwise, the feature is selected. To express feature subset selection problem using chromosomes, a number generator that returns 1 or 0 randomly can be used to generate chromosomes.

2.3 Fitness function

In the feature subset selection system, a fitness function is used to evaluate how "good" a feature and the level of association of that feature with the fault values [9][11].

To give a score to each set bit (recall that one set bit refers to a feature) of chromosome, Pearson Correlation Coefficient (PCC) is adopted. PCC is a statistical measure of interdependence of two or more random variables. The PCC correlation coefficient is defined as r

$$r = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_i (x_i - \overline{x})^2}\sqrt{\sum_i (y_i - \overline{y})^2}} \tag{1}$$

The expression can be simplified into the following form:

$$r = \frac{S_{xy}}{S_x \bullet S_y} \tag{2}$$

where X (x_1, x_2, ….x_i)represents features and Y (y_1, y_2, ….y_i) represents the fault. Sxy is the standard deviation of $X \bullet Y$, and Sx and Sy refers to the standard deviation of X and Y respectively. This simplifies the implementation as only a few arrays to store intermediate results.

As the magnitude of r is needed for scoring of a particular bit of the chromosome, and its value can be negative, the absolute of r is returned as the score of that feature. To filter out low r value as zero association, a confidence level is introduced so that only r values above the confidence level are considered in the total score. Therefore, set bits that have a score less than the confidence level is given a score of 0 and the bit is reset to 0. In this way, the results only consist of selected features whose r values are higher than the confidence level.

2.4 Selection using roulette wheel

Simple reproduction allocates offspring strings using a roulette wheel with slots sized according to fitness [12]. This is one way of choosing members from the population of chromosomes that is proportional to their fitness. Parents are selected according to their fitness. The better the fitness of the chromosome, the higher the chance it will be selected. However, the fittest member is not guaranteed to go to the next generation.

Imagine a roulette wheel where all chromosomes in the population are placed according to fitness, as in Figure 1 below. The wheel is "spun" and the marble fallsinto one of the slots. The larger the slot, the higher the chance it will be selected.

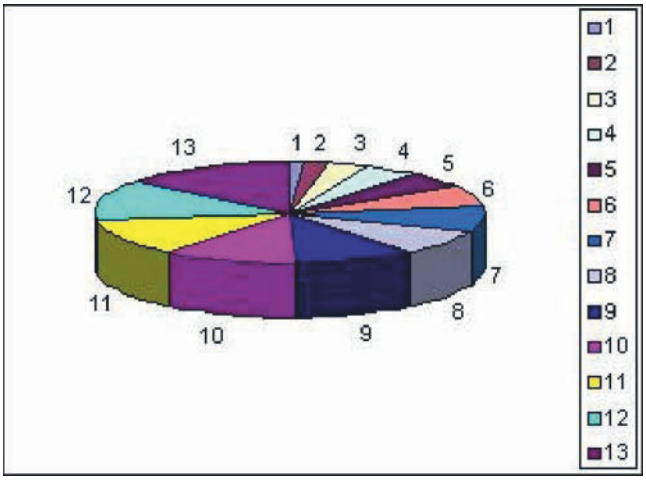


Figure 1. A roulette wheel with thirteen chromosomes

2.5 Crossover and mutation

One point crossover is implemented in our method by picking one position in the chromosomes at random and then exchanging the parent chromosomes to form children chromosomes as in Figure 2. Mutation is much easier to implement. To mutate a certain bit, we generate a number randomly and compare the number with the mutation rate. If the random number is smaller than the mutation rate, the bit is mutated; otherwise, the bit is maintained and next bit is considered. The mutation is shown in Figure 3.

Parent 1	1011001010110010
Parent 2	1111110000011001

Crossover point

Offspring 1	1011110000011001
Offspring 2	1111001010110010

Figure 2. Single-point cross over in GAs

Mutation Point	
Offspring 1	1011**1**10000011001
Mutated Offspring 1	1011**0**10000011001

Figure 3. Mutation in GAs

3 CASE STUDY

A case study is carried out to verify the usability of the method. An application related to high speed machining tool condition is selected for the experiment. Tool condition is an important factor in the high speed machining process. Tool wear and tool failure may result in a loss in surface finish and dimensional accuracy of the finished parts, and even possible damage to the work piece and machine [13].

3.1 Experimental set up and features extraction

Most CNC machines are not able to detect machining tool's wear-out in an on-line manner. The cutting force signal is instead used to establish usable models due to its high sensitivity to tool wear, low noise, and good measurement accuracy [14]. Our experiment about the tool condition is shown in Figure 4.

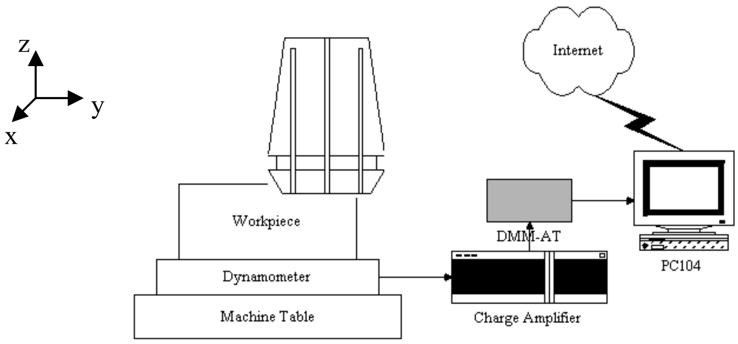


Figure 4. Tool Conditions Experiment in Machinery

The cutting force along the y-direction of the machining path was measured by a dynamometer in the form of charges, and converted to voltages by the Kistler charge amplifier. The voltage signal was sampled by the PCI 1200 board at 2000 Hz and directly streamed to the hard disk of a computer. The flank wear of each individual tooth of the cutting tool was measured using an Olympus microscope. Details of the experiment set up and feature extraction methodologies have been reported elsewhere [15]. Four experiments were used for our study and the cutting parameters are summarized in Table 2.

Table 2

Cutting Experiments

Test No	Spindle Speed (rpm)	Feed Rate (mm/min)	Depth of Cut (mm)	Insert Number
1	1000	100	1	4
2	1000	200	1	2
3	1200	150	1	2
4	1000	100	1	2

From these experiments, 16 main features are extracted and summarized in Table 3. The 16 features that have been shown to be effective for tool wear-out monitoring form the scope of the feature subset selection in this paper.

Table 3

Feature Extraction Methodologies

No	Feature	Notation
1	Residual Error	Re
2	First Order Differencing	Fod
3	Second Order Differencing	sod
4	Maximum Force Level	fm
5	Total Amplitude of Cutting Force	fa
6	Combined Incremental Force Changes	df
7	Amplitude Ratio	ra
8	Standard Deviation of the Force Components in Tool Breakage Zone	fstd
9	Sum of the Squares of Residual Errors	sre
10	Peak Rate of Cutting Forces	kpr
11	Total Harmonic Power	thp
12	Average Force	Fa
13	Variable Force	vf
14	Standard Deviation	std
15	Skew	skew
16	Kurtosis	kts

3.2 Feature subset selection

Feature subset selection can be achieved following the steps mentioned in Section 2. Although the feature subset selection results do not depend on the input parameters, the performance of the feature subset selection depends largely on the setting of the starting parameters. Hence, an investigation of how each parameter affects the performance of the system was conducted.

3.2.1 Parameters used for GAs

The input parameters and their representation are summarized in Table 4.

Table 4

Parameters and their representation in GAs

Parameter	**Representation**
Population Size	The size of the population for every generation
Crossover points	The number of crossover points to be defined when performing crossover
Crossover rate	The chance of crossover being applied to a chromosome
Mutation rate	The chance of a chromosome being mutated
Confidence level	A limiting value on the required association *r* value.
Maximum Number of generation	The maximum number of generations the GA is allowed to evolve to. Also the terminating condition of the algorithm.
Number of test sets	The amount of data sets used for the calculation of *r*.

Since crossover rate and mutation rate only elevate the randomness of the system, their influence on the performance of the subset selection system is unnecessary [16]. Hence, these two parameters study are omitted. To investigate the influence of the other five input parameters on the performance of the feature subset selection, experiments were performed for 30 runs for each parameter. From our experiments, the parameters that have insignificant effect on the system's performance are number of crossover points, maximum number of generations, the number of test sets. These parameters produced similar results and a wrong setting may increase the run-time, hence leading to the degradation of the system performance. As for other two parameters (population size and confidence level), optimal values improved the system's accuracy. The optimal parameter setting from our experiments are shown in Table 5.

Table 5

Optimal parameters setting used in GAs

Parameter	**Value**
Population Size	20
Crossover points	1
Crossover rate	0.7
Mutation rate	0.001
Confidence level	0.9
Maximum Number of generation	200
Number of test sets	2000

3.2.2 Feature selection results using GAs

Experiments were conducted with different numbers of selected features at beginning. We found that the least number of features to build association model with fault is four. If feature numbers are less than four in the feature subset selection, the correlation between feature subset and fault can not be well established in our case. Fifty runs of each experiment data set on feature subset selection with four features were conducted according to the optimal parameter setting shown in Table 5. The feature selection results are shown in the following Table 6.

Table 6

Experiment Result

		Data Set (50 runs)			
		Test No 1	Test No 2	Test No 3	Test No 4
Features	re				
	fod				
	sod				
	fm	49	47	47	44
	fa	50	48	47	47
	df				
	ra				
	fstd				
	sre				
	kpr				
	thp				
	Fa	49	48	48	47
	vf				
	std	47	49	49	40
	skew				48
	kts				47

The results were labeled with the feature notation and for each data set, the number of times a feature is identified as relevant to tool wear is counted.

From the results, the features that were discovered to be most significant against tool wear were {fm, fa, Fa, std}. Each of these features has an r (fitness) value of at least 0.95. The average time taken for 50 runs of each of the data sets was about 230 seconds.

3.3 Evaluation of the feature selection

The selected features were used to build a multiple regression model to evaluate selected feature subset. Four-variable regressions are used to generate the relation model between wear-out and the features. Wear-out prediction is calculated based on the regression model. Four runs of experiments with different compound features are conducted and the results are discussed below.

The experiment condition and parameters of Test No. 1 are described in Table 2. Totally 65,536 sets of data are captured in Test No. 1. We used half of the data to conduct the training and another half to test our prediction model. Through data processing, the experiment data are separated into two sets, odd numbers of the data (1, 3, 5, ...) are used for training and the even numbers (2, 4, 6, ...) of data for the testing.

We first used the selected feature subset from GAs {fm, fa, Fa, std}, The regression equation obtained from the training data is as in equation (3). The prediction and real tool wear-out comparison diagram is shown in Figure 5 (a).

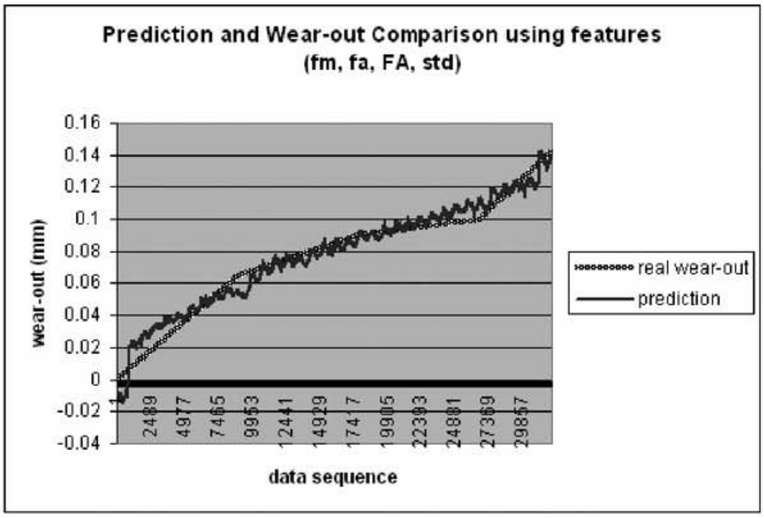
$$W_{ear\text{-}out} = -0.187+0.238f_m - 0.004f_a - 0.0129F_a + 0.054std \quad (3)$$

We second used some randomly selected feature subset {fm, fod, sod, vf}, The regression equation obtained from the training data is as in equation (4). The prediction and real tool wear-out comparison diagram is shown in Figure 5 (b).

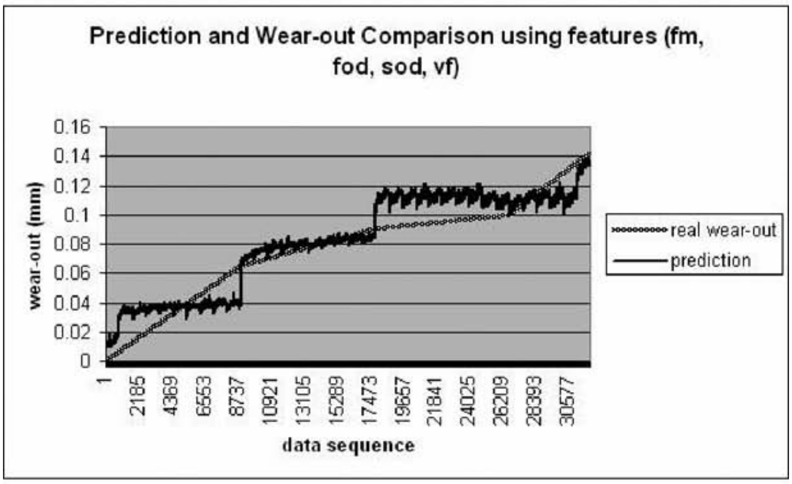
$$W_{ear\text{-}out} = -0.127+0.220\,f_m+0.543\text{ fod} - 0.0353\text{ sod} + 7.795\text{ vf} \quad (4)$$

We third used some other randomly selected feature subset {sre, std, thp, vf}, The regression equation obtained from the training data is as in equation (5). The prediction and real tool wear-out comparison diagram is shown in Figure 5 (c).

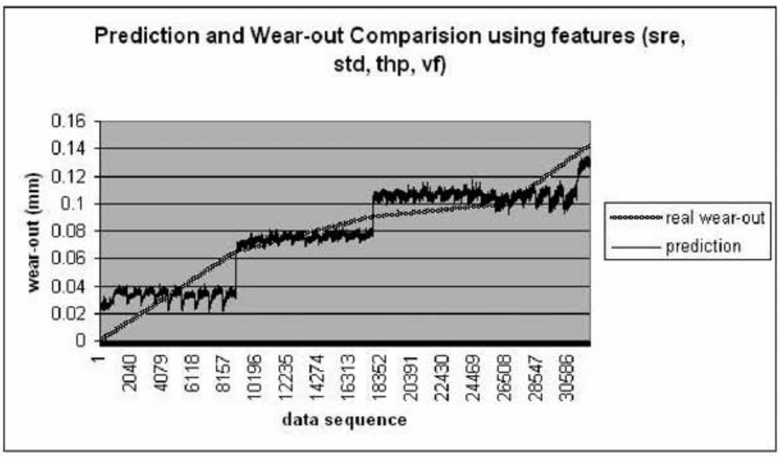
$$W_{ear\text{-}out} = -0.104 + 6.439\text{sre}+0.199\text{ std}+0.003\text{ thp} - 0.578\text{ vf} \quad (5)$$



(a)



(b)



(c)

Figure 5. Prediction and Wear-out Comparison

The prediction and wear-out comparison using different feature subset are shown in Figure 5. The prediction mean error, average relative error and estimation of variance between the prediction and real wear-out for different feature subsets are summarized in Table 7. The selected feature subset from GAs gives the best prediction result for tool wear-out.

Table 7

Comparison results of prediction accuracy for different features subsets

	features (fm, fa, FA, std)	features (fm, fod, sod, vf)	features (sre, std, thp, vf)
Prediction mean error (mm)	0.005	0.012	0.011
Average relative error	11.89%	22.70%	22.50%
Estimation of variance	0.237	0.445	0.425

To conduct further study, we used other three sets of experiments data to evaluate the feature subset selection. R^2 of the multiple regression models is used for evaluation in this study. R^2 is usually used to test the degree of association between output variable Y and all the explanatory variables X_i. R^2 lies between 0 and 1. When R^2 equals 1, the fitted regression line explains 100 percent of the variation in Y. On the other hand, when R^2 equals 0, the model does not explain any of the variation in Y. The closer R^2 to 1, the better the fit of the model [17]. The comparison results of R^2 from different features subsets are shown in Table 8. The R^2 value from selected four features {fm, fa, Fa, std} are much closer to 1 than other randomly selected features and this verifies that the selected features from GAs has the most correlation to tool's wear-out.

Table 8

Comparison results of R^2 for different feature subsets

Test No	R^2 for 16 features	R^2 for four selected features (fm, fa, FA, std)	R^2 for four randomly selected features (fm, fod, sod, vf)	R^2 for four randomly selected features (sre, std, thp, vf)
1	0.979	0.964	0.843	0.873
2	0.971	0.953	0.80	0.846
3	0.985	0.969	0.856	0.887
4	0.972	0.950	0.82	0.841

Our experiments also found that use original 16 features to build multiple regression models takes about 5 times of computational time than to build the models with four selected features. Therefore, use the four selected features subset from GAs has the advantage to implement on-line prediction as it saves computational time without loss of accuracy.

4 CONCLUSION

In the present work, a modified wrapper-based multi-criterion approach using genetic algorithms has been designed for solving feature discovery problem. The best compound feature subset by GAs is verified by using multiple regression models and used to build fault prediction models. The method has been tested experimentally with the prediction of tool wear-out on a high speed milling machine. An acceptable prediction of the tool wear-out has been achieved, which can be viewed as a demonstration of the usability of feature selection method in the manufacturing environment.

5 REFERENCES

[1] R. Kohavi, and G. John (1997). Wrappers for Feature Subset Selection, Artificial Intelligence, 97:273-324

[2] M.Raymer, W. Punch, E. Goodman, Li. Kuhn and A. Jain (2000). Dimensionality Reduction Using Genetic Algorithm. IEEE Transactions on Evolutionary Computation.

[3] David Beasley, David R Bull and Ralph R Martin (1993). An Overview of Genetic Algorithm Fundamentals, 15(2):58-69.

[4] M. Kudo and J. Sklansky (2000). Comparison of algorithms that select features for pattern classifiers. Pattern Reconitin, 33(1): 25-41.

[5] W. Sielecki and J. Sklansky (1989). A note on genetic algorithms for large scale on feature selection. Pattern Recognition Letters, 10:335-347.

[6] Lac, H. C. and D. A. Stacey (2005). Feature subset selection via multi-objective genetic algorithm. Proceedings of the International Joint Conference on Neural Networks 2005, 31 July-4 Aug. 2005, Montreal, Que., Canada, IEEE.

[7] Yang, J. and V. Honavar (1998). Feature subset selection using a genetic algorithm. IEEE Intelligent Systems & Their Applications 13(2): 44-49.

[8] Vafaie, H. and K. De Jong (1992) Genetic algorithms as a tool for feature selection in machine learning. Proceedings. Fourth International Conference on Tools with Artificial Intelligence, 10-13 Nov. 1992, Arlington, VA, USA.

[9] Chakraborty, B. (2002). Genetic algorithm with fuzzy operators for feature subset selection IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences E85-A(9): 2089-2092.

[10] Kuah Tong Sing, "Feature Subset Selection Using Genetic Algorithms", Final Year Project Report, Nanyang Technological University, 2004

[11] Peterson, M. R., M. L. Raymer, et al. (2005). Balanced accuracy for feature subset selection with genetic algorithms. The 2005 IEEE Congress on Evolutionary Computation, 2-5 Sept. 2005, Edinburgh, Scotland, UK, IEEE.

[12] Marek Obitko, Roulette Wheel Selection, http://cs.felk .cvut.cz/`xobitko/ga/selection.html, 1998

[13] Kuo, R. J. and Cohen, P. H. (1999) Multi-sensor integration for on-line tool wear estimation through radial basis function networks and fuzzy neural network. Neural Networks 12, pp. 355-370.

[14] Altintas, Y., et al, (1989) In process Detection of Tool Failure in Milling Using Cutting Force Models. J. Engng Ind., Vol. 111, pp. 149-157.

[15] Dong Jianfei, et al, (2004) Bayesian Support Vector Classification for Maching Process Monitoring. Fourth International ICSC Symposium on Engineering Intelligent Systems, Portugal.

[16] A.E. Eiben, R. Hinterding and Z. Michalewicz (1999) Paramer control in evolutionary algorithm. IEEE Trans. On Evolutionary Computation, 3(2):124-141.

[17] Damodar, N. Gujarati. (1995) "Basic Econometrics", Third Edition, McGRAW-HILL INTERNATIONAL EDITIONS.

EFFECT OF INFORMATION ON THE RISK OF UNEXPECTED FAILURE

D Chaschin

ABB Performance Service Centres. Automation Technologies Division, ABB Australia

Abstract: Condition monitoring (CM,) as part of the Reliability Based Maintenance, is getting more and more popular. With the right systematic approach, this method is proven to increase the productivity of the most manufacturing processes. There is no doubt that when CM helps to find the developing fault and prevents the unexpected failure, it proves to be a good investment. What happens, if the fault doesn't exist? Is this information still useful? How do we estimate the value of this information? How can we use this information in development of the maintenance strategy? This paper has some answers to these questions. Although all these issues are important for any technical objects, the author is investigating the relations between the information and reliability in electrical machines and in induction motors in particular, because they are currently the most popular drives in the industry. The examples are given to demonstrate the approach.

Key Words: Reliability, Condition Monitoring, Information

1 INTRODUCTION

The failure of any mechanism is the result of many possible causes. Unexpected failure of any mechanism is always a result of one reason - insufficient information. This statement determines the easiest solution for reliability improvement: instead of trying to prevent the failure we can try to predict it and undertake all necessary action to prevent or minimise the losses that the failure may cause. This strategy has two important advantages: it is achievable and it is cheaper.

This strategy also has limitations. In most cases gathering all off the information about the mechanism is impossible, sometimes it is very expensive. The attempt to gather the exhaustive information about your equipment can simply deplete your maintenance funds. To assess the amount of information we need, we have to know how to assess the effect of information on the risk of unexpected failure.

Information, by Claude Shannon's definition, is the measure of uncertainty. This definition automatically gives us the answer to the question about the use of information. The value of information can be assessed by the degree of the uncertainty reduction.

The most important question about the equipment involved in the production process is: "How long is it going to last?" Three corner stones are needed to answer this question:

- adequate model, that contains all the parameters affecting the process;
- knowing that all the parameters, used in the model have correct values;
- knowing that the parameters are not changing in time or knowing how they are changing.

Unfortunately there is very limited number of mechanisms that have accurate answers to those questions. That is why we have to acknowledge that the time to failure is a random value.

The other question worth considering is: What equipment we would call more reliable:

- the one that operates 100 hours before failure?
- the one that operates 1000 hours before failure?
- the one that operates 10000 hours before failure?

Subconsciously most of us would pick the last one, but in fact all of the above equipment is equally unreliable if it fails UNEXPECTEDLY. This brings us back to the importance of the information as a tool to prevent the unexpected failure.

2 RELATION BETWEEN RELIABILITY AND INFORMATION

"The development of this program is towards the control of reliability through an analysis of the factors that affect reliability and provide a system of actions to improve low reliability levels when they exist." [1]

"Developments in the automation and electronics in last decades caused the problem of the reliability of the complex multi-element systems which still remains unresolved." [2]

"The problem of reliability is becoming more important with every year". [3]

"…Reliability is such a critical parameter of most modern engineering product" [4]

All the above quotes have one common feature: they are underlining the importance of the reliability problem for manufacturing. But they also have a difference: the first three quoted books were published in 1961 and the last one was published in 2004.

Why the reliability remains the problem for the last 50 years? The following diagram may help to find the answer.

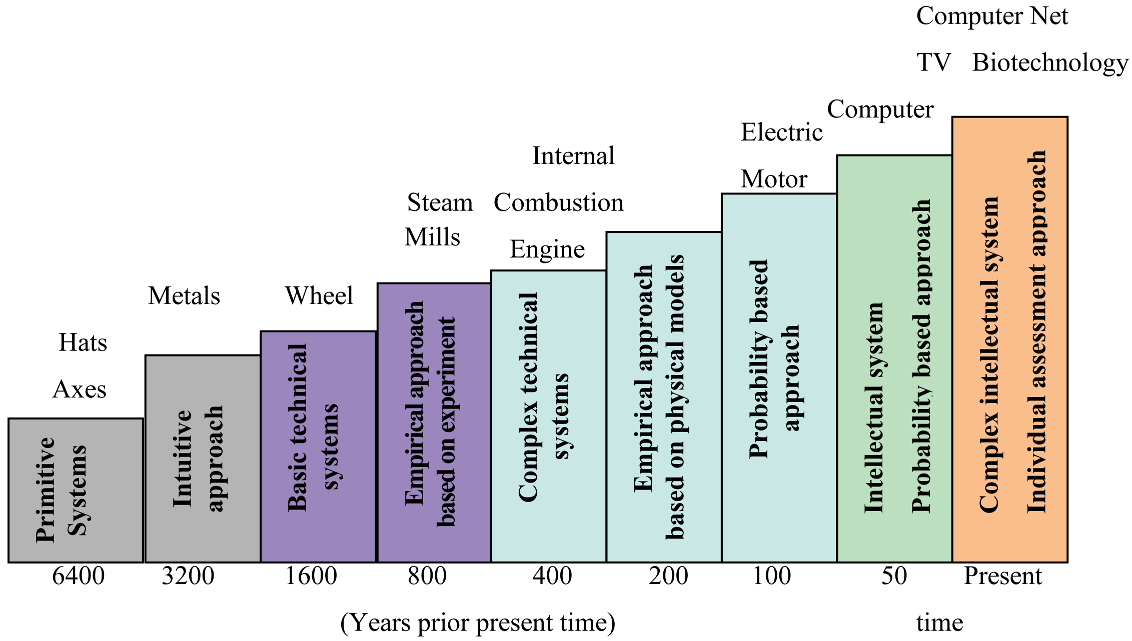


Figure 1 Relation between the complexity of the systems and approach to reliability.

The diagram shows that as long as there were primitive systems and very little knowledge about them, there was an intuitive approach to the reliability by "go/no go" tests. The failure could happen any time with the same probability.

Development of basic technical system such as the carriage, mill, hat etc required some basic knowledge that was mainly gained from experience with some simple empirical modelling. That was better than nothing, but didn't do much for the improvement of reliability.

The next period contains complex technical systems that required complex physical models. The calculations were more accurate, the modelling was more sophisticated and as a result the reliability could be controlled to a certain degree. But the knowledge was still not refined and to achieve higher reliability the reserve factors were widely used. It was working well before the complex mobile systems were developed.

It is understood that one can improve the reliability of the shaft simply by increasing its diameter. But it will increase its weight as well. And if you can afford it for some stationary equipment, it is always in contradiction with weight restriction of mobile systems, which always mean to be lighter. This contradiction initiated the probability-based approach to reliability problems. This approach is in fact a compromise between the level of knowledge we have about the system and reserve factor we can afford. So instead of a simple guess we could have an educated guess.

As a result of the technical evolution, more complex technical systems were involved in operation without the methods of accurate control of their reliability. In 1952 reliability problems in electronic equipment forced the US Department of Defence and the electronics industry to jointly set up the Advisory Group on Reliability of Electronic Equipment. The recommendations of this group later became the US Military Standard 781, Reliability Qualification and Production Approval Tests. This can be considered as the beginning of the systematic approach to reliability problems.

In 1965 the principles of the above standards were developed further and lead to the integration of the reliability engineering approach into the common engineering approach. These new principles were put into the new standard MIL-STD-785-Reliability Programs for Systems and Equipment. This approach was used by NASA for their space program.

The next level in reliability development was reached in 1980s, when a new Japanese approach took the western world by surprise. This approach emphasised the total quality management role in reliability improvement and it lead to considerable increase in reliability with reduction in price.

The outcome of the above historical issue is that the level of reliability development was lagging behind the development of technical systems. Only in industries such as defence and space exploration, where reliability is absolutely crucial, the new methods were implemented. Other industries considered such approaches too expensive and preferred to stick to the good old "repair when fail" methods. That is why the cost of maintenance of the equipment sometimes reaches 40% of the manufacturing expenses. Considering that manufacturing expenses are counted in hundreds of billions, this 40% extra cost really reaching cosmic proportions. It is time to understand that the lack of information causes low reliability. The lack of information is not necessary caused by our negligence; in most cases we simply don't know that this particular factor will affect the reliability of our system. No matter how well designed the system could be, there will be always some factors that are not considered in the design. Here is simple example from electric motors. European manufacturer receives an order from Africa for several motors. The motors were specially designed for protection from excessive heat and moisture. After some time of operation the motors started to fail. The investigation showed that local ants liked the insulating materials used in the motors. This example proves that it is impossible to provide high reliability only by good design. Although it is a very important part of the equipment life cycle, there are also manufacturing and operations that will inevitably bring their corrections to the initial design.

Let us see how the above principles reflect on the electrical machines. Induction motors are the most popular and reliable electrical machine.

3 ANALYSIS OF THE INDUCTION MOTOR FAILURE

What is causing the uncertainty in the life expectancy of the induction motor? The diagram on the figure 2 would help us to understand it. It shows the most important parameters that affect induction motor characteristics through its lifetime and also subject to variation. As you can see, the first stage of the motor life cycle is its design and the outcome of the design can be affected by the specifications such as required power, voltage, character of the load, ambient temperature etc. The more detailed the provided specifications can be, the more adequacy the motor design will have. If the motor with open air circuit is working in an aggressive environment, like high concentration of SO_3, then no matter how well it is maintained, it will fail quickly. Another factor that affects the design is the model that the designer is using. How accurate is the model? How accurate is the loss calculation, because this will determine the efficiency of the motor and the temperature rise. How accurate are the starting characteristics? This will determine how quick the starting process will be, and the starting process is the most stressful for the most of the motors.

Next step in the life cycle is manufacturing, and the outcome of this process depends on two factors as well. There are materials that are usually supplied by subcontractors and the manufacturing process itself, which includes coil manufacturing and insertion, lamination stamping and core assembly etc. All the factors shown as "Manufacturing" can vary and, as a result, change the motor characteristics. Unfortunately there are not many parameters in the manufacturing process, deviation from which would improve the reliability of the electric motor. The problem is that the electric motor was invented more than 150 years ago and most of the manufacturers already reached maximum possible outcome from the motors of a particular design. It requires new materials or new energy transformation principles to be discovered in order to improve the electric motors any further. That is why variations of the manufacturing parameters from their designated values can only degrade the motor characteristics including the reliability.

The transport and set up steps have the same effect. All the impacts of the transportation as well as misalignment during the setup can severely damage the bearings and cause unexpected failure.

A lot of variations happen to the motor operation conditions. Only the most critical are shown on the diagram below. It is important to understand, when assessing the factors affecting motor reliability in operation, that it is not only the motor features that affect its reliability, but also supply, load and ambient conditions. It is also important that most of the factors listed in the diagram are mutually connected. For example quality of the core manufacturing will affect the flux density, which in turn will affect the starting characteristics and the temperature of the motor. What is interesting, is that the load conditions will determine wether the temperature will rise or fall.

As it can be seen from the diagram and also confirmed by experience, there are a lot of factors which affect the motor life and make it very uncertain. The diagram also shows that neither of the life cycle steps is able to improve the reliability. Only a complex approach including all steps of the motor life cycle will bring positive results.

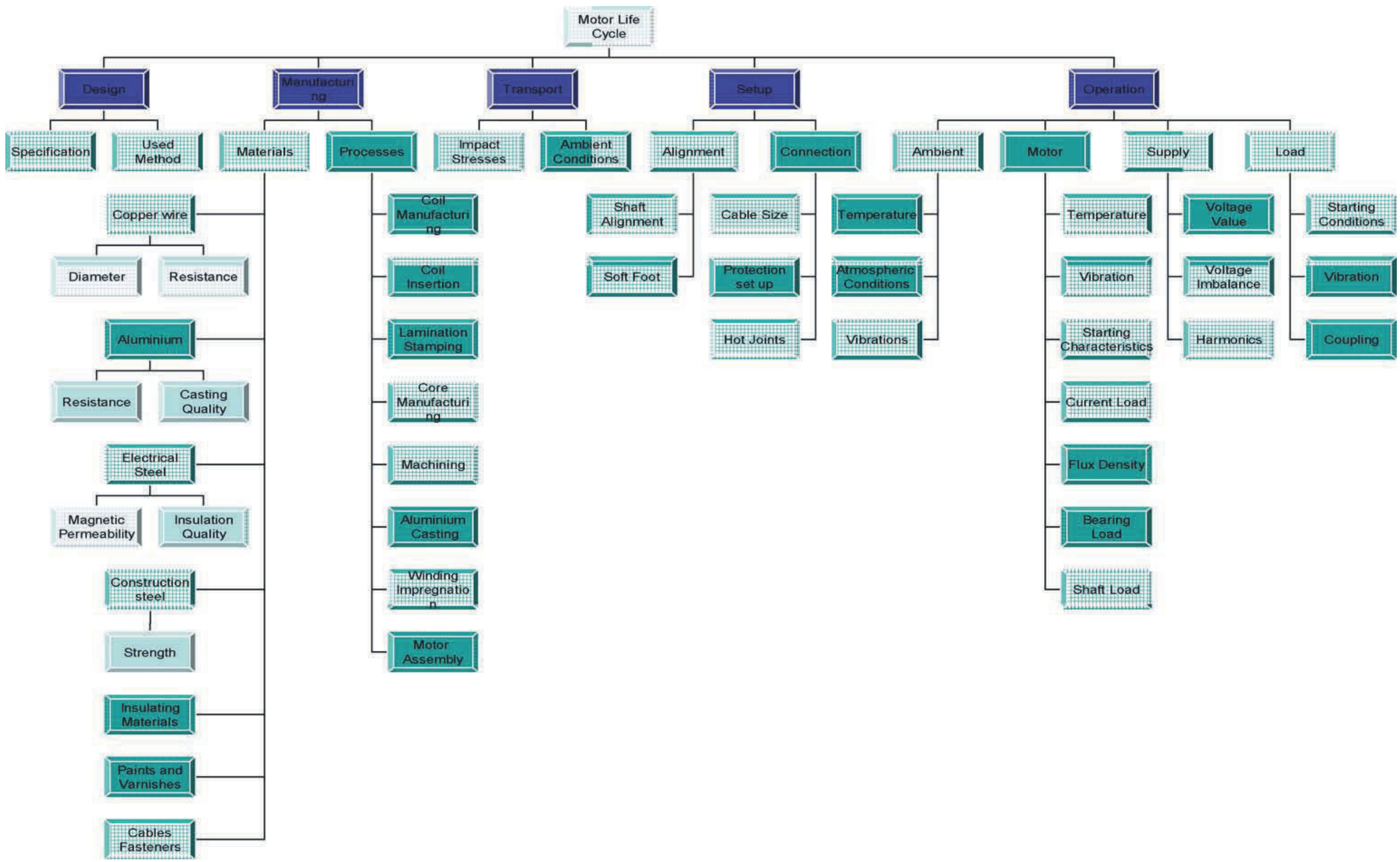


Figure 2. Electric Motor life cycle.

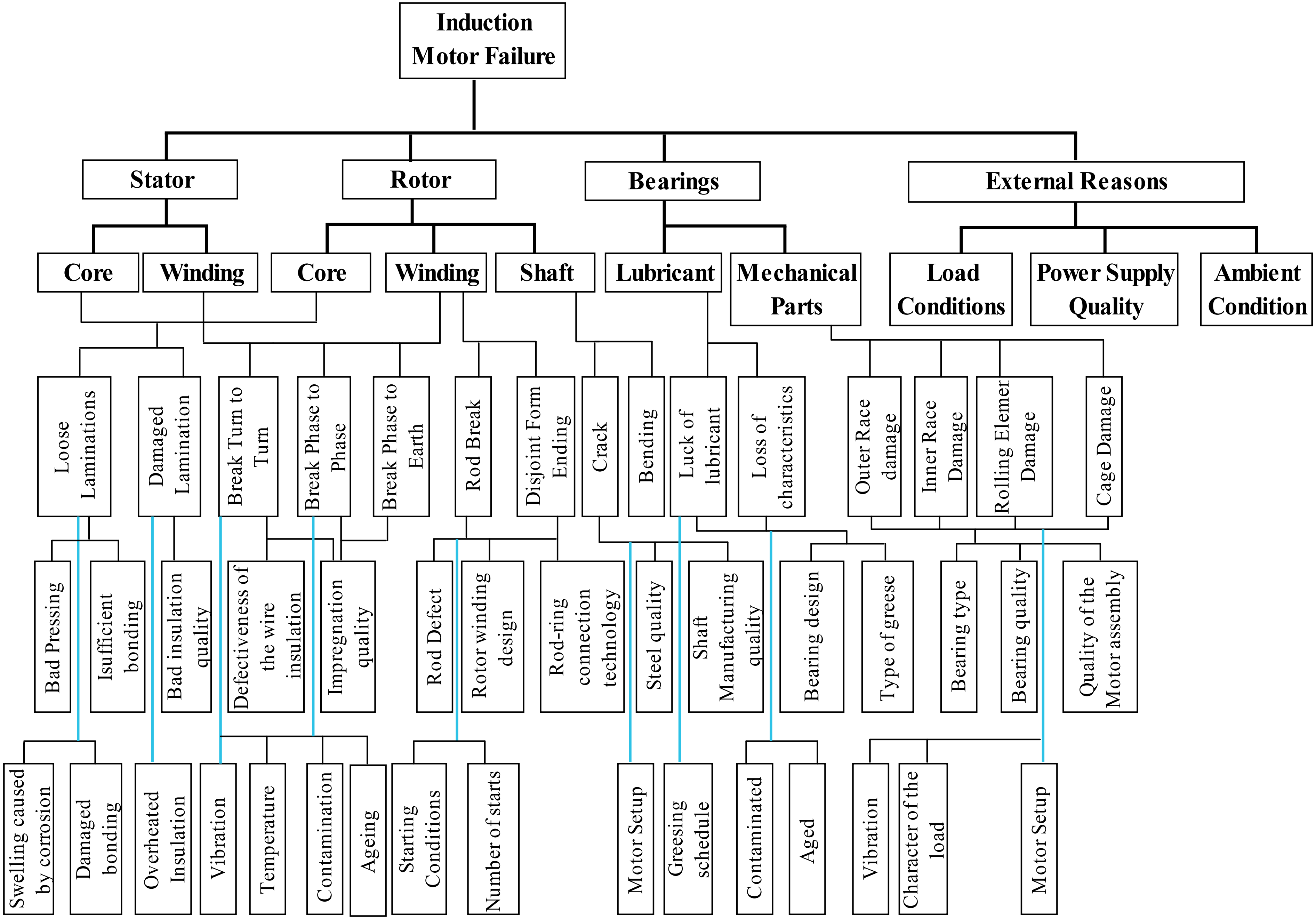
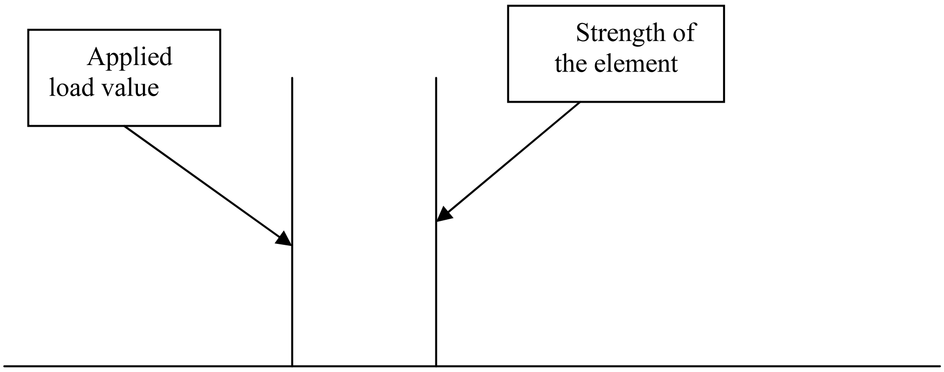


Figure 3. Failure diagram of the Induction Motor

Diagram on the Fig 3 shows possible failure types of the induction motors. There are many common blocks in those diagrams. The significance of the diagram on the Fig 3 is that it shows the connections between particular life cycle parameters and type of faults. Analysis of these diagrams shows the possible ways of the prevention of the unexpected failures. The diagrams show the most important elements of the Induction motor reliability system and connections between them. Let us have a look at this problem from the statistical point of view.

4 STATISTICAL PRESENTATION OF THE RELIABILITY

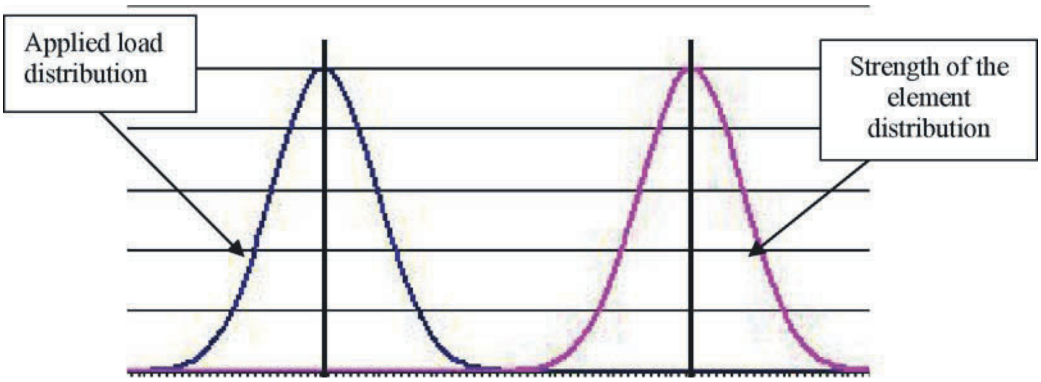
It is well known that the failure of the technical system happens when the applied load exceeds the strength of the element it is applied on. If all the principles of absolute reliability that were laid down in the introduction to this paper would be fulfilled, then we would have the following model:



Load-Strength

Figure 4 Load-Strength discrete values

When we have the adequate models of the load and strength and know that all the parameters are accurate, we will have the discrete values of the load and strength and it will provide absolute 100% reliable product. We have to change the word “know” in the above sentence to the words “have enough information” to bring it closer to the subject of this article. And this is the problem. As we discussed before, most of the time, by different reasons, we don’t have enough information. In this case we are facing another option, presented on Figure 5 below.



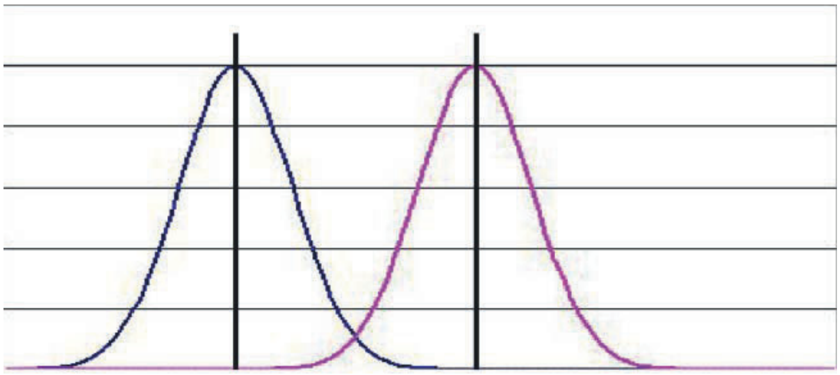
Load-Strength

Figure 5 Load Strength distributed values

The distributed values simply represent the lack of knowledge. It means that both load and strength are affected by some parameters that are included neither in the model nor in the measured values. In other words we don’t have enough information to accurately determine load and strength. Imagine how accurate the results of your electric current calculation would be if you would measure only resistance of the circuit, but not the voltage.

But in this case the reliability is again very close to 100% because the distribution curves are not interfering, which means that even the highest possible load doesn’t reach the value of the lowest possible strength. Obviously the mean value of the strength in this case should be substantially higher than the mean value of the load, represented by the according straight lines.

Now if we decided to reduce the strength of the element, for example reduce the shaft diameter, then we will have the situation presented on the Figure 6, the most common case in engineering reliability.



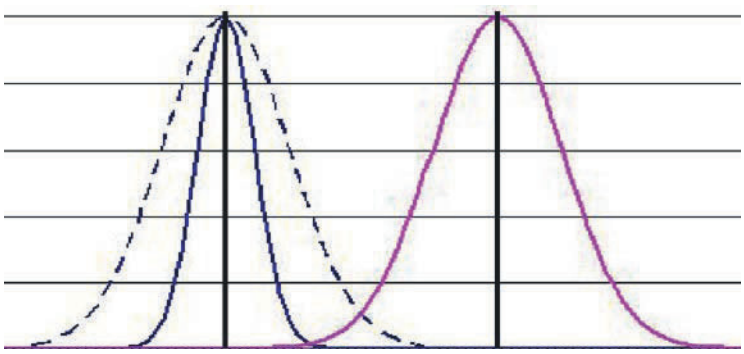
Load-Strength

Figure 6 Load-Strength interfering distributions

In this case because of the lowered mean value of strength we have the interference of the distributions. This means that in some cases the applied load will be higher than the strength of the element and there will be a chance that the element will fail.

And here we have our critical moment. There are two ways of fixing the problem:

- increase the mean value of the element strength, so we will have the case presented on the Figure 5
- reduce the variation of the load or strength or both and have the new case presented on the figure 7.



Load-Strength

Figure 7 Load strength reduced load variation

As you can see, the mean values of the load and strength didn't change, but the probability of the load exceeding the strength practically disappeared.

Now the only question is how to reduce the variation of the load or of any parameter for that matter. This is where the information comes on stage. Modifying the design model by including new parameters that were not considered before, analysing the manufacturing process trying to find the parameter that creates the most uncertainty in the final product and offcourse monitoring the operating conditions of the equipment will reduce the variations of load and strength and by doing that, increase the reliability of the equipment. That is where diagrams 2 and 3 can be very useful when talking about electric motors.

From this chapter we can conclude that it is not the mean values that create the problems with reliability. The mean values are usually analysed all the way through the life cycle. It is the variations that create the problems and it is the variations that should be considered as the main target in the reliability improvement process.

5 THE METHOD OF THE SENSITIVITY ASSESSMENT

As it was written in the introduction, and concluded from the previous chapter, it is important to know which parameter is the most influential in respect to reduction of the uncertainly. The method of the sensitivity assessment can help in this case. As it was mentioned before, it is the variation of the parameters not their designated values, that affect the reliability most. Mathematically we can use the following equation:

$$Y_{jo} = f(x_1, x_2, \ldots x_i, \ldots x_n) \tag{1}$$

Here Y_{jo} - designed value of the j^{th} characteristic from the upper level of the diagram Fig.2;

x_i - i^{th} designed parameter from the lower level of the same diagram,

The actual value of the characteristic after the motor was manufactured and installed can be described as:

$$Y_{jo} + \Delta Y_{j\Sigma} = f(x_1 + \Delta x_1, \ldots x_i + \Delta x_i, \ldots x_n + \Delta x_n) \tag{2}$$

Here $\Delta Y_{j\Sigma}$ - deviation of the j^{th} characteristic caused by the deviations Δx_1 of the parameters from the lower level.

There are following assumptions made:

the deviations of the characteristics and parameters are small, i.e.

$$\Delta Y_{j\Sigma} << Y_{jo} \text{ and } \Delta x_i << x_i \tag{3}$$

- because of (3) the relations between the deviations of characteristics and deviations of the technological parameters are linear or can be presented as linear;
- the distribution of the technological parameters is close to the Gauss distribution.

Then, using Taylor series we can write:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2!} f''(x_0)(x - x_0)^2 + \frac{1}{3!} f^{(3)}(x_0)(x - x_0)^3 + \ldots + \frac{1}{n!} f^{(n)}(x_0)(x - x_0)^n \ldots \tag{4}$$

Based on the assumption about the linearity of the function $f(x)$, we can ignore all the derivatives higher than first order and rewrite (4) as:

$$f(x) - f(x_0) = f'(x_0)(x - x_0) \tag{5}$$

With some alterations we can rewrite (5) as follows:

$$\frac{f(x) - f(x_0)}{f(x_0)} = f'(x_0) \frac{x - x_0}{x_0} \frac{x_0}{f(x_0)} \tag{6}$$

Here $\frac{f(x) - f(x_0)}{f(x_0)}$ - relative function variation;

$\frac{x - x_0}{x_0}$ - relative argument variation.

For several variables (6) looks like this:

$$\frac{\Delta f(x)}{f(x_0)} = \sum_1^n \left[\frac{\partial f(x_1, x_2, \cdots x_i, \cdots x_n)}{\partial x_i} \bullet \frac{x_i}{f(x_1, x_2, \cdots x_i, \cdots x_n)} \right] \frac{\Delta x_i}{x_i} \tag{7}$$

Or in terms of the characteristics presented on the diagrams 2 and 3:

$$\frac{\Delta Y_{j\Sigma}}{Y_{jo}} = \sum_1^n \left[\frac{\partial f(x_1, x_2, \cdots x_i, \cdots x_n)}{\partial x_i} \bullet \frac{x_i}{f(x_1, x_2, \cdots x_i, \cdots x_n)} \right] \frac{\Delta x_i}{x_i} \tag{8}$$

The coefficients in square brackets are called sensitivity coefficients C_{ij} .

This equation allows investigation of the sensitivity of variation of each parameter from upper level to the variation of each parameter from lower level.

If we consider changing only one parameter x_i , then the derivatives of the other parameters of the sum in square brackets will be equal to zero and (8) can be presented as:

$$\frac{\Delta Y_{ji}}{Y_{jo}} = C_{ji} \frac{\Delta x_i}{x_i} \tag{9}$$

To explain this process in more details, let us go through the example.

6 EFFECT OF THE INFORMATION ABOUT THE BELT TENSION ON THE RISK OF THE BEARING FAILURE

To analyse this issue we need to know the model that connects the load on the bearing with its life expectancy. The most popular model is the one developed by Palmgren and Lundberg and it is as follows:

$$L_h = K_\lambda K_t \frac{10^6}{60n} \left(\frac{C_h}{Q} \right)^m \tag{10}$$

Here Lh – basic rating life, hours; n – motor RPM; Ch – basic dynamic load rating, N;

Q – dynamic load, applied to the bearing during the operation time, N;

$K_{\lambda,} Kt$ - corrective coefficients depending on the lubrication conditions and temperature, m – bearing type factor equal to 3 for ball bearings and 10/3 for roller bearings.

. It is known that the time before failure for rolling bearings is described by Weibull distribution:

$$F(t) = \begin{cases} 1 - e^{-\frac{(t-a)^k}{b}} & if (t \geq a) \\ 0 & \text{if t< a} \end{cases} \tag{11}$$

Here a – location parameter; b – scale parameter and k – shape parameter.

For ball bearings shape parameter k = 1.5. Then scale parameter for 90% probability can be calculated as:

$$b = \frac{t_\gamma}{\sqrt[k]{-\ln(\gamma/100)}} = \frac{t_{90}}{\sqrt[1.5]{-\ln(90/100)}} = 4.48 t_{90} = 4.48 L_{10} \qquad (12)$$

Here L_{10} - basic rating life of the bearing used in the most of the bearing hand books.

Now the Reliability function of the ball bearing can be presented as:

$$R(t) = \exp\left[-\left(\frac{t}{4.48 L_{10}}\right)^{1.5}\right] \qquad (13)$$

Here t- time of operation, hours; L_{10} – basic rating life of the ball bearing, hours.

Now we can have a look at the example:

Induction motor is driving a fan through the pulley and set of v-belts.

The DE bearing is 6305. The other parameters are as follows:

Basic dynamic load rating	C = 22500 N
Mean value of the dynamic load	P = 1000 N
Rotation speed	n = 1500 RPM
Expected belt tension force	Pb = 400 N; force variation from 100 to 700 N
Expected magnetic pull force	Pm = 150 N; force variation from 0 to 300 N

All other components (fan load, contact angle of the bearing, temperature, lubrication characteristics etc.) are presumed stable.

1	Variation of the belt tension force	700-100 = 600 N
2	Estimation of the standard deviation	600/6 = 100 N
3	Variation of the magnetic pull:	300-0 = 300 N
4	Estimation of standard deviation	300/6 = 50 N
5	Resultant variation	$\sqrt{100^2 + 50^2}$ = 111.8 N
6	Estimated mean life expectancy according to Lundberg/Palmgren bearing equation (10)	$L_{10} = K_\lambda K t \frac{10^6}{60 n} \cdot \left(\frac{C}{P}\right)^3$ = 1*1*10^6/60/1500*(22500/1000)^3 = 12656 h
7	Standard deviation in per load units	111.8/1000 = 0.1118
8	Sensitivity of the time to failure to the variation of the load according to (8)	K_Q = -3
9	Deviation of the time to failure because of the load variation	3*0.1118 = 0.3354
10	Deviation of the time to failure in hours:	12656*0.3354 = 4245 h
11	Shape parameter of the Weibull distribution for	σ_L / M_L=0.3354 is approximately 1.85
12	Scale parameter for γ=90 using (12),	b=3.375L_{10}

13 Probability of failure before 10000 hours according to (13):

$$R(t) = 1 - \exp\left[-\left(\frac{t}{3.375L_{10}}\right)^{1.85}\right] = 0.066$$

If we monitored the belt tension and keep it at 400N, the variation of the Pb would be 0.

Substitute 0 in the above calculation will give us probability of failure:

R(T) = 0.054

This example shows that information about only one parameter can reduce the risk of failure by more than 20%. Note that the mean values of the load levels are remaining the same for both cases. We also can name such parameters, related to the bearing reliability, as vibration, eccentricity of the air gap, rotor imbalance. Modern testing methods allow measuring these parameters as well as the belt tension and this additional information will reduce the risk of the unexpected failure even more.

Presented paper explains the reasons of uncertainty in estimation of life expectancy in rotating machines and shows that additional information can reduce this uncertainty. Presented diagrams and the method of sensitivity assessment help to establish the most efficient information sources.

7 REFERENCES

[1] FAA Industry Reliability Program. Federal Aviation Agency. November 7, 1961 p.1

[2] Bazovsky I. (1961) Reliability Theory and Practice. New Jersey, Prentice-Hall, Inc.

[3] Looyd D, Lipov M (1961) Reliability: Management, Methods, and Mathematics. NJ, Prentis-Hall, Inc.

[4] Summerville N (2004) Basic Reliability: Introduction to Reliability Engineering. Author House.

FACILITIES MANAGEMENT ACTION AGENDA:THE AUSTRALIAN GOVERNMENT AND INDUSTRY WORKING TOGETHER FOR IMPROVEMENT

S Ballesty [a bei]

[a] *Rider Hunt Terotech, Level 5, 41 McLaren Street, North Sydney, New South Wales 2060, Australia*
[b] *Facility Management Association of Australia [1]*

Abstract: This paper provides an overview of the Australian Government's Facilities Management (FM) Action Agenda as announced in 2004 as a key policy plank designed to facilitate growth of the FM industry. The resulting consultation with industry leaders has seen the creation and release in April 2005 of the FM Action Agenda's strategic plan entitled *'Managing the Built Environment'*. This framework, representing a collaboration between the Australian Government, public and private sector stakeholders and Facility Management Association of Australia (FMA Australia) and other allied bodies, sets out to achieve the vision of a more "*…productive and sustainable built environment…*" through improved innovation, education and standards. The 36 month implementation phase is now underway and will take a multi-pronged approach to enhancing the recognition of the FM industry and removing impediments to its growth with a 20 point action plan. Additional information is available at www.fmactionagenda.com.au

Keywords: facilities management, recognition, innovation, education and training, regulatory reform, sustainability, built environment

1 INTRODUCTION

Facilities Management (FM) is "a business practice that optimises people, process, assets and the work environment to support delivery of the organisation's business objectives" (Glossary of Facility Management Terms, 1998).

According to Access Economics' 2004 statistical profile of the industry, FM is a significant contributor to the Australian economy. The combined direct and indirect contribution of the FM industry in 2002-03 was A$12.2 billion of value added, A$12.4 billion in GDP terms, and (full time equivalent) employment of 172,000 persons. The combined contributions represented 1.80%, 1.65% and 2.10% share of the corresponding Australian GDP and employment totals.

In 2004 the Australian Government announced their support of the FM Action Agenda as a key policy plank designed to facilitate growth of the FM industry. The resulting consultation with industry leaders and collaboration between the Australian Government, public and private sector stakeholders, and industry bodies saw the release in April 2005 of the FM Action Agenda's strategic plan entitled '*Managing the Built Environment'*.

The FM Action Agenda sets out to achieve the vision of a more "*…productive and sustainable built environment…*" through improved innovation, education and standards.

This paper provides an overview of the process thus far, its strategic directions and the current implementation phase.

2 ABOUT ACTION AGENDAS

Action Agendas are a key policy plank of the Australian Government under the Department of Industry, Tourism and Resources (DITR). Action Agendas represent an opportunity for collaboration between Industry and Government, with both working together to identify impediments to growth, harness sustainable competitive advantage and to facilitate growth of specific sectors.

Action Agendas also provide industry direction for a whole-of-government approach on key issues such as innovation, investment, the environment, market access, regional development, education and training, workplace relations and regulatory reforms.

In October 2003 the Facility Management Association of Australia (FMA Australia) successfully made application to the DITR. On 19th January 2004 the federal Industry Minister, Hon. Ian Macfarlane MP announced the development of a Facilities Management (FM) Action Agenda.

There are currently 36 Action Agendas at various stages (4 under development, 15 being implemented through the adoption of recommendations, and 17 completed, as at December 2005). The industries covered by these Action Agendas accounted for about a third of Australia's gross domestic product (GDP) and employment, about 46 per cent of Australia's export earnings, and contributed more than 75 per cent of Australia's business investment in research and development (R&D).

Action Agendas provide a flexible model for industry to consider in developing sectoral priorities, and planning for the future in partnership with government.

3 STRATEGY FOR CHANGE

"It seems appropriate that in the Year of the Built Environment, we will create the opportunity for this growing industry .... it is industry commitment combined with the support of the Australian Government that Action Agendas require to be a success. This sector is positively positioning itself to deliver improved productivity on a whole range of work fronts."

Hon. Ian Macfarlane MP, Minister for the Department of Industry, Tourism and Resources (January, 2004)

3.1 Providing Leadership

Announced on the 19th January 2004, the FM Action Agenda saw FMA Australia work with DITR representatives to form a Strategic Industry Leaders Group (SILG) for the strategic phase. Nominations from a broad spectrum of industry participants were received and during February - March 2004 ministerial appointments were made of individuals drawn from the following organisations, representing a mix of private and public sector facility users, FM service providers and consultants.

- AEH Property
- ANZ Banking Group
- Coles Myer
- Connell Mott MacDonald
- Department of Industry, Tourism and Resources
- Hames Sharley
- CSIRO Corporate Property
- Investa Property Group
- Multiplex Facilities Management
- Rider Hunt
- Resolve FM
- Spotless Group
- Stockland Corporation
- Telstra Corporation
- Tungsten Group

The SILG comprised 15 industry leaders appointed by the Hon Ian Macfarlane MP, Minister for Industry, Tourism and Resources, and chaired by John McCarthy.

3.2 Formulating our Agenda

The SILG met several times to develop the focus areas, it was recognised from early on that a lack of an agreed definition of FM and recognition of its significant contribution was a core issue. Without identity and recognition any group of partitioners will struggle to attract quality candidates, access resources, achieve professionalism, develop standardisation and ultimately sustainability.

Practical objectives identified by the SILG included;

- Recognition by all levels of Government and corporate sectors of the value of FM in terms of environmental performance, economic contribution, immigration, innovation and research, education and training, etc;
- Provision of a national focus for the FM innovation and compliance to maximise efficiencies and effectiveness;
- Ensuring that procurement and management decisions are made on the basis of whole-life costs, cultural fit and not solely short term financial criteria;
- Providing leadership, building capability and sharing knowledge;
- Improving partnering and collaboration between service providers and contractors;
- Creating clearer career paths and specific tertiary education courses;
- Support of professional, dynamic, moral and ethical attitudes;
- Improved productivity and sustainability on a whole range of fronts;
- Justification of the business case for FM investment; and
- Value for money.

"Whole-of-government support through the acceptance of our Action Agenda proposal is an exciting and critical step forward in the development of the facility management industry."

Steven Gladwin, Chairman, FMA Australia (January, 2004)

3.3 Providing Focus

At the SILG's first meeting on 31st March 2004 five (5) core themes identified and agreed, including;

- Recognition
- Innovation
- Education and Training
- Regulatory Reform
- Sustainability

Working groups were established to develop the themes.

Each of the five (5) Working Groups was chaired by a member of the SILG. Many had more than one SILG member actively participating. In all the Working Groups involved some 50 volunteers from some of Australia's leading organisations, who gave not only of their time and energy, but also covered their own expenses. The task of the Working Groups was to identify, according to their respective themes, key issues affecting competitiveness and future growth for the industry.

The findings for each Working Group, along with recommendations were presented to the SILG for consideration as part of developing the FM Action Agenda.

The SILG met on the 26th May and 21st July 2004 to review the progress of the Working Groups. The SILG also met on the 13th October 2004 to agree and prioritise actions, and to consider draft action agenda report. The drafting process continued until the 1st December 2004 when the SILG signed off the FM Action Agenda report for submission to the Minister and to consider the implementation strategy.

When launched at Sydney Opera House by Parliamentary Secretary, the Hon. Warren Entsch, MP, on 28th April 2005 the FM Action Agenda was acknowledged as "… the best ever Action Agenda …" with a clear and concise vision of the future.

3.4 Statistics for the Future

FM is not formally recognised in Government policy or an identified industry by the Australian Bureau of Statistics (ABS) or the Department of Immigration, Migration and Indigenous Affairs (DIMIA). This is reflective of the prevailing lack of recognition of the distinctive nature and importance of FM.

As part of the FM Action Agenda strategic phase Access Economic were commissioned to conduct an independent and credible analysis of the Facilities Management industry size and contribution to the Australian economy and compile a statistical profile.

Good quality statistics are essential to any industry, and especially to an emerging one industry such as Facilities Management, which is at a critical stage in its development. Statistics can provide information on production and employment trends, inter-industry relationships, market opportunities, potential threats and a range of other matters that point to the industry's future and influence decision-making.

The Access Economics 2004 study has provided a starting point for collecting statistics on the industry. In the short-term, the industry can build upon this methodology to update current statistics. Over the longer term the industry could look to the Australian Bureau of Statistics to provide improved data.

For the purposes of the Access Economics report, FM activities are concerned only with servicing of buildings. The services include services provided by both FM management and FM service suppliers such as tradesmen and cleaners. The business-specific operating costs of occupants of the facilities were excluded. This means that FM costs consist mainly of building-related repairs and maintenance, servicing and energy for lifts and air conditioning, gardening / landscaping, cleaning, lighting, security, and management fees, whether actually paid to FM firms or imputed (ie. carried out 'in-house' by building owners or tenants). Occupant costs for all other energy uses are excluded, as are any costs of production and processing within a facility.

The analysis was not restricted to businesses formally designated as FM firms. It was intended to include all activities that have the potential to be undertaken by FM firms, regardless of whether or not these activities may be undertaken 'in-house'. Access Economics found boundaries of the industry are not defined precisely. For example, while all servicing of buildings used by educational and hospital facilities have the potential to be covered by FM contacts, there is some arbitrariness in deciding where to draw the boundaries on the sizes of office buildings, retail centres and multi-unit residential buildings that are to be included.

As a general rule, where Access Economics had to make a judgement on these matters, it generally erred on the side of understating the size of the FM industry.

Table 1

Facilities Management Contribution to the Australian Economy (2002-03)

	Value Added ($mil)	GDP ($mil)	Employment ('000)
Direct management component	1,958	1,982	16
Direct service supplier component	6,583	6,662	119
Total direct	8,451	8,644	135
Total indirect	3,689	3,753	37
Combined direct and indirect	12,230	12,397	172
Combined share of Australian economy	1.80%	1.65%	2.10%

Source: Access Economics (2004)

4 MANAGING THE BUILT ENVIRONMENT

4.1 FM in Australia

The Facilities Management industry evolved during the 1980's as a consequence of more widespread outsourcing by business of activities, such as the management and maintenance of buildings. The industry plays a significant role in planning and managing the built environment throughout its life cycle, which has three linked stages: facility creation, operation and disposal.

The industry's primary function is providing strategic management advice on the efficient operation of buildings and reducing operational life cycle costs. The industry is also responsible for ensuring that services are delivered in a way which contributes to the productivity and profitability of building occupants, as well as improved returns for owners and investors.

The range of services provided include traditional building services such as repairs and maintenance, security and cleaning, as well as more highly technical services requiring skilled personnel. The FM industry contributes $12.4 billion to Australia's gross domestic product (GDP). This is equivalent to 1.65 per cent of Australia's GDP. It includes a direct GDP contribution of $8.6 billion or around 1.15 per cent of Australia's GDP, and an indirect contribution of a further $3.8 billion or about 0.50 per cent of Australia's GDP. Total employment by the FM industry was 172,000 persons in 2002-03. This is equivalent to about 2.1 per cent of Australia's workforce.

The FM Action Agenda aims to develop a strategic framework for the growth of a sustainable and internationally competitive Australian Facilities Management sector.

The SILG examined five main strategic areas which constitute the main areas of focus for the FM Action Agenda. The areas identified were: facilities management in the Australian economy, innovation, education and training, regulatory reform, and sustainability. The SILG also examined issues associated with implementing the FM Action Agenda and raising awareness about the facilities management industry more generally.

4.2 Innovation

Success in achieving growth in the FM industry will depend on the promotion of an innovative culture and bringing innovations into the market place.

The actions proposed promote the benefits of innovation through greater industry collaboration and research and development, and to highlight the contribution that facilities management makes to workplace productivity. An FM Innovation Forum will be established to facilitate the exchange of information and ideas, and a web portal designed and built to disseminate this information, 31st May 2006 saw the launch of www.fmactionagenda.com.au

4.3 Education and Training

Education and training is essential to creating a skilled FM workforce which is recognised by clients as professional and capable of delivering a valuable service.

It has been proposed that industry work with the vocational education and training (VET) sector to support the provision of VET, and with universities to increase the availability of facilities management courses at the undergraduate level. Greater collaboration between the VET and higher education sectors is also proposed, along with an awareness campaign to increase recognition of Facilities Management as a career of choice by school leavers.

4.4 Regulatory Reform

An awareness campaign has been proposed focusing on three important regulatory areas asbestos, emergency evacuation plans and cooling towers with five further identified areas (dangerous goods; fall prevention; confined spaces; plant safety, and electrical inspections testing and residual current devices) to follow. It is recommended that the industry undertake initiatives to further the cause of regulatory reform in these and other areas. Industry should seek to participate in the various regulatory consultative mechanisms to promote conformity of requirements and contribute to key areas impacting on the efficiency of Facility Managers.

4.5 Sustainability

The FM industry is in a strong position to influence decisions made by government and corporate entities to lower environmental impacts through business practices.

The industry should actively promote its contribution to the sustainability challenge for business and the community. Further, FM should be aware of the role it can play in key industry and government forums on promoting sustainability. The new FM Action Agenda web portal should also be used to disseminate information and develop a business case model highlighting the costs and benefits of sustainable practices.

4.6 Implementation

The SILG proposed that an Implementation Group be formed comprising leading people drawn from within the FM industry to follow through on the 20 actions incorporated in the FM Action Agenda strategic plan.

Such an Implementation Group would need to work closely with all industry stakeholders and meet regularly for a period of up to three years after the launch of the FM Action Agenda. The SILG also proposed an awareness campaign designed to communicate clear messages about the contribution and multiplier that FM represents to Australia's Built Environment.

The key objective of the FM Action Agenda is to support the future growth and sustainability of Australia's facilities management industry by successfully developing and promoting its competitive advantages.

5 IDEAS INTO ACTIONS

FM Action Agenda industry vision being that Facilities Management should be; "the foremost contributor to a productive and sustainable built environment through excellent and innovative management of facility services."

The following is the 20 point action plan as developed by the FM Action Agenda and as published in *'Managing the Built Environment'* (2005), and now being followed though in the Implementation Phase;

5.1 Facilities Management in the Australian Economy

Action 1	Annually update data on the facilities management industry's contribution to the Australian economy
Action 2	Consult with the Australian Bureau of Statistics with a view to obtaining improved data for the facilities management industry.

5.2 Innovation

Action 3	Establish a Facilities Management Innovation Forum for the facilities management industry.
Action 4	Establish a web portal for the facilities management industry to disseminate information and provide feedback on innovation, education and training, regulatory issues, and sustainability.
Action 5	Promote the benefits of innovation and encourage greater industry collaboration and investment in research and development.
Action 6	Highlight the contribution that facilities management makes to productivity by establishing performance measures and 'best practice' benchmarks, incorporating client input, with a specific focus on improving workplace productivity.

5.3 Education and Training

Action 7	Work with the Construction and Property Services Industry Skills Council to support the provision of vocational education and training (VET) for the facilities management industry.
Action 8	Work with Australian universities to increase the availability of facilities management courses at the undergraduate level.
Action 9	Develop an innovative project that will encourage greater collaboration between the VET and higher education sectors in the provision of facilities management education.
Action 10	Undertake an awareness campaign directed at those who influence career choice – including school counsellors, career advisors, educators, parents and students – to increase understanding of facilities management as a career.

5.4 Regulatory Reform

Action 11	Promote awareness within the facilities management industry of the regulatory requirements applying in three selected areas (asbestos; emergency evacuation plans; and cooling towers).
Action 12	Identify the regulatory requirements for a further five identified areas (dangerous goods; fall prevention; confined spaces; plant safety; electrical inspection and testing and residual current devices).
Action 13	Promote conformity of regulatory requirements and contribute to key areas of regulatory reform.

5.5 Sustainability

Action 14	Promote the role of facilities management in responding to increased demand for corporate accountability associated with sustainability performance.
Action 15	Promote the role of the facilities management industry in key industry and government forums addressing sustainability.
Action 16	Use the data web portal proposed in Action 4 to disseminate sustainability information.
Action 17	Develop a 'business case' model that highlights the costs and benefits of embracing sustainable practices in the use and management of materials; energy; water; waste; and indoor environmental quality; with a particular focus on workplace productivity.

5.6 Implementation

Action 18	Establish an Implementation Group comprising leaders of the facilities management industry to manage the implementation of the Action Agenda.
Action 19	Develop a communications strategy to increase recognition of the contribution of facilities management as an industry.
Action 20	Work with relevant government departments and agencies to identify opportunities to raise awareness of the facilities management industry and to consult on relevant government policies and programs.

5.7 FM Exemplar Project

Additionally the SILG resolved to commission an FM Exemplar Project for the Implementation Phase as a demonstrator of best practice to act as a base for industry technology transfer and up-skilling. Several facilities at various stages of their life cycle were considered, however Sydney Opera House was ultimately selected on 13th October 2004 based on its ability to cover a range of the 20 action areas and its iconic status.

6 IMPLEMENTATION PHASE BEGINS

"The FM Action Agenda puts our profession on the map and provides all industry participants and stakeholders with the opportunity contribute to delivering a productive and sustainable built environment through improved innovation, education and standards."

Stephen Ballesty, Chairman, FMA Australia (September, 2005)

Following the launch of the FM Action Agenda publication entitled *"Managing the Built Environment",* which is the output of the strategic phase conducted throughout 2004 in consultation with industry, again saw FMA Australia and DITR representatives working form a Strategic Industry Leaders Implementation Group (SILIG). Again nominations from a broad spectrum of industry participants were received and during August 2005 ministerial appointments were made of individuals drawn from the following organisations, representing a mix of private and public sector facility users, FM service providers and consultants.

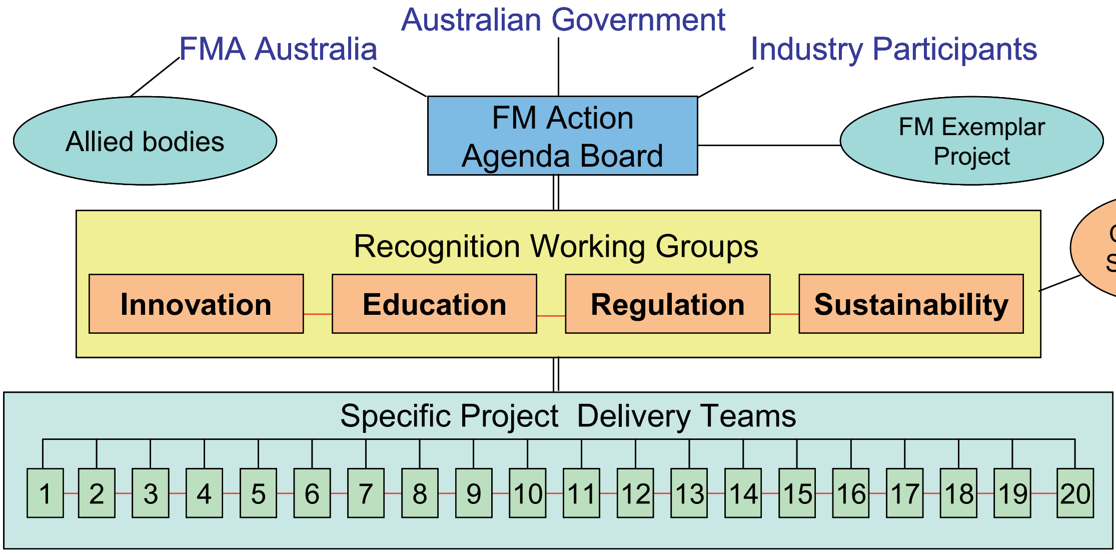
- AEH Group
- Australian Department of Parliamentary Services
- Balance Corporate Property (from July 2006)
- Investa Property Group
- Macquarie Bank
- Mirvac Group (until November 2005)
- Multiplex Facilities Management
- Myer
- Queensland Department of Public Works
- Resolve FM (until March 2006)
- Rider Hunt Terotech
- Spotless Group (until February 2006)
- Transfield Services
- Tungsten Group
- University of New England

This voluntary body comprising 14 industry leaders with an *ex officio* member from the Department of Industry, Tourism and Resources, again chaired by John McCarthy has now been officially designated the FM Action Agenda Implementation Board. The Board at their first meeting reviewed the FM Action Agenda strategy and the 20 action items for the 36 month Implementation phase, and resolved;

- its structure and responsibilities around the focus areas of Innovation, Education & Training, Regulatory Reform and Sustainability;
- that FMA Australia would act as facilitation and provide support for the process;
- allocated specific responsibilities for the formation of Recognition Working Groups for each of the four focus areas; the formulation of a Communications Strategy and liaison with the FM Exemplar Project – Sydney Opera House in order to progress the 20 action items;
- to appoint a Board Secretary and seek funding for a scoping study to identify the requirements for Project Delivery Teams; and
- a meeting and reporting cycle for the future.

Figure 1

Implementation Phase Structure (diagrammatic only)



Source: Stephen Ballesty

"This is a defining moment for the Facilities Management and I am delighted to be part of such a dedicated team which represents a mix of experience gained during the strategic phase and new talent that would contribute to important industry transformation".

John McCarthy, Chairman, FM Action Agenda Implementation Board (September, 2005)

The 36 month implementation phase is now underway and will take a multi-pronged approach to developing the FM industry and removing impediments to growth with a 20 point action plan across the following platforms:

- *Innovation* – Improved appreciation of facility life cycles, and greater understanding of the key drivers of workplace productivity, and the improved application of information technology.
- *Education and Training* – Improved access to dedicated FM education and training opportunities and creation clear career pathways into the profession.
- *Regulation Reform* – Explore opportunities to harmonise cross jurisdictional regulatory compliance requirements that have an efficiency impact on FM.
- *Sustainability* – Improved utilization of existing knowledge and the development of tools and opportunities to improve the environmental performance of facilities.

As of January 2006, the DITR Parliamentary Secretary, Bob Baldwin, MP, has assumed the oversight responsibility for the FM Action Agenda. The Implementation Board of 12 with John McCarthy as Chairman is making good progress on the four (4) Recognition Working Groups, Communications Strategy and FM Exemplar Project. The first FM Action Agenda annual report is due in August 2006.

7 DEMONSTRATING BEST PRACTICE

The *FM Exemplar Project: Sydney Opera House* was launched on 28th April 2005 as a CRC-Construction Innovation project with a team including the CSIRO, the University of Sydney, Sydney Opera House, Transfield Services and Rider Hunt Terotech. The project is due for completion by the end of 2006.

Sydney Opera House has a design life of 250 years primarily as a 'performing arts centre', "providing first class venues, facilities and services that support our artistic and business aspirations." according to Sydney Opera House Trust Annual Report (2004). However, at the same time is acknowledged as an 'architectural masterpiece' and an 'asset of heritage significance', being regarded as having 'iconic' status and value for 20th century architecture and contributes to the tourism value of Sydney. The facility comprises 7 theatres, 37 plant rooms, 12 lifts and over 1000 rooms. Sydney Opera House employs 300 full-time staff and a further 300 part-time staff, delivering over 2500 performances per annum. Sydney Opera House provides some unique FM challenges.

Sydney Opera House currently has an annual FM expenditure budget of approximately A$19 million pa – across a wide range of activities from minor maintenance through to significant refurbishments and upgrades. Sydney Opera House's FM Department has recently initiated a series of FM enhancements to the way they procure and implement FM services.

Additionally there is a planned future upgrade of the Opera Hall (estimated cost of A$400 million over the 2006-2008), ongoing acoustic enhancements, fire and hydraulic upgrades and maintenance, and architectural refurbishment works.

The FM Exemplar Project combines three (3) research streams dealing with *Digital Modelling*, *Services Procurement* and *Benchmarking* as a whole and then develop collaboration between them. It aims to achieve innovative FM strategies and models that will showcase improved FM performance and promote best practice.

Sydney Opera House has acknowledged the challenges they face with their existing documentation management. A digital backbone of Building Information Model (BIM); or accurate three dimensional (3D) geometry with attributes that define the detail of building elements, and relationships to other objects; will be critical for future FM activities. The *Digital Modelling* theme will develop an appropriate BIM strategy and investigate the potential of state-of-the-art information systems to enable a future integrated platform to support FM activities and processes.

The *Services Procurement* theme focuses on the maintenance and cleaning services procurement, especially with reference to in-house versus outsourcing methodologies, in the context of the user requirements, prevailing market and the service provision reliability and risk. Various procurement methods and a multi-criteria assessment approach for supporting decision making are to be investigated.

The *Benchmarking* theme focuses on the strategic asset management of iconic facilities with similar functionalities, with a bias for the performing art centres. Critical success factors in the functional areas of FM are to be identified against organisational objectives and key performance indicators in order to better monitor, control and improve FM with respect to performance targets seeking participation from international comparators. Collaboration between Services Procurement and Benchmarking includes sharing of the benchmarking data and utility of key performance indicators to support innovative procurement strategies.

Further it is envisaged that a significant component of the project budget will be expended on Australia-wide workshops, industry brochures including '*How To'* guides, conference and journal publications, and course development through higher education and TAFE institutions. In addition, Sydney Opera House will be able develop best practice benchmarks not only to justify the iconic facility status but to demonstrate industry leadership in FM.

8 CONCLUSIONS

The Facilities Management Action Agenda via its strategic plan released in April 2005, entitled '*Managing the Built Environment'* represents a collaboration between the Australian Government, public and private sector stakeholders, FMA Australia and other allied bodies. There is now a real commitment to achieve the FM Action Agenda's vision of a more "…*productive and sustainable built environment*…" through improved innovation, education and standards.

Arguably, the most significant challenge the FM industry faces is lack of recognition, not only of its distinct and relatively recent character, but of FM as a significant economic contributor and multiplier. Without identity and recognition any group of partitioners will struggle to attract quality candidates, access resources, achieve professionalism, develop standardisation and ultimately sustainability. The FM Action Agenda is a real opportunity to address such issues.

The 36 month implementation phase is now underway and will take a multi-pronged approach to enhancing the recognition of the FM industry and removing impediments to its growth with a 20 point action plan across the following platforms:

- *Innovation* – Improved appreciation of facility life cycles, and greater understanding of the key drivers of workplace productivity, and the improved application of information technology.
- *Education and Training* – Improved access to dedicated FM education and training opportunities and creation clear career pathways into the profession.
- *Regulatory Reform* – Explore opportunities to harmonise cross jurisdictional regulatory compliance requirements that have an efficiency impact on FM.
- *Sustainability* – Improved utilization of existing knowledge and the development of tools and opportunities to improve the environmental performance of facilities.

In addition, the FM Exemplar Project represents an excellent opportunity to identify best practice leverage off the iconic nature of Sydney Opera House's international and national profile to identify and develop best practice within the FM industry. The innovations and methodologies delivered by this Project should be implemental across the FM industry at the strategic, management and operational levels, with clear training and educational benefits leading to improved service procurement and delivery.

The FM Action Agenda makes a clear case for Facilities Management as an enabler of business change and community benefit.

9 REFERENCES

[1] Access Economics (2004), *Facilities Management – A Statistical Profile of the Industry*

[2] Facilities Management Action Agenda (2005), *Managing the Built Environment, Commonwealth of Australia*, Australian Government, Department of Industry, Tourism and Resources, Canberra

[3] Facility Management Association of Australia (2001), *Glossary of Facility Management Terms,* Standards Australia HB261

[4] Facility Management Association of Australia (1998), *FM Research Report 1: Facility Management in Australia*, Arthur Andersen

[5] Facilities Management Exemplar Project: Sydney Opera House (2005), *Digital Modelling, Services Procurement and Benchmarking research*; CSIRO, University of Sydney, Sydney Opera House, Transfield Services Australia and Rider Hunt Terotech

[6] Sydney Opera House Trust Annual Report (2004)

[7] www.construction-innovation.info

[8] www.fma.com.au

[9] www.fmactionagenda.com.au

10 ACKNOWLEDGEMENTS

The author wishes acknowledges the Australian Government's Department of Industry, Tourism and Resources' representatives and staff for their support of the FM Action Agenda, and the contributions of the members and their organisations of the FM Action Agenda's voluntary Strategic Industry Leaders Group, Implementation Board and the associated Recognition Working Groups, Project Delivery Teams and FM Exemplar Project Team and the Facility Management Association of Australia.

---

[i] *S Ballesty* [c d]

Steven Gladwin, Stephen Ballesty, Kerry Lodge, Alan McGrath, Sue Pridmore, Greg Tenbrink and Richard Mayes (Board as at 2003) and Stephen Betros (until Sept. 2004) and Karen Hill (CEO from Dec. 2004)

[c] *FM Action Agenda [(2)] Strategic Industry Leaders Group: John McCarthy (Chairman), David McCarthy, Stephen Ballesty, Ian Fox, Steve Gladwin, Jane Hamilton, George Harley, Chris Hunt, Ross Johnston, Andrew Junor, Warren Kerr, Jon McCormick, Ross Murchie, John Phillis and Sue Pridmore*

[d] *FM Action Agenda [(3)] Implementation Board: John McCarthy (Chairman), Stephen Ballesty (Deputy Chairman), Kevin Dickinson, Steven Gladwin, Chris Hunt (until May 2006), Ross Johnston (until Feb. 2006), Chris Luscombe (Sept. – Nov. 2005), Karen Lyon-Reid, Jon McCormick, John Nakkan, Naomi Nielsen, Sue Pridmore, Michael Silman, Tony Staveley (from July 2006) and George Spink*

[e] *FM Exemplar Project [(4)] CRC for Construction Innovation: Peter Scuderi, Kylie Legge, and Marcello Tonelli (Mar to May 2006); CSIRO: Lan Ding, Robin Drogemuller, John Mitchell and Hans Schevers; University of Sydney: David Leifer, Dirk Schwede, Jeremy Wu (until April 2006), Janet Henriksen, Kat Martindale (Feb to April 2006); Sydney Opera House: Paul Akhurst; Transfield Services: George Spink; and Rider Hunt Terotech: Stephen Ballesty, Jason Morris (April to Nov. 2005) and Ankit Shah (from May 2006); and RMIT: Ron Wakefield.*

[1] *Initial advocacy in October 2003, and ongoing support and secretariat to the present*

[2] *Strategic Phase Leadership: January 2004 to March 2005*

[3] *Implementation Phase Leadership: April 2005 to present*

[4] *FM Exemplar Project Team: April 2005 to present*

AN INTEGRATED COLLABORATIVE APPROACH FOR FM– SYDNEY OPERA HOUSE FM EXEMPLAR

Stephen Ballesty [a b]

[a] *Rider Hunt Terotech*

Level 5, 41 McLaren Street, North Sydney, New South Wales 2060, Australia

[b] *Cooperative Research Centre for Construction Innovation*

Level 9, L Block, QUT Gardens Point, 2 George Street, Brisbane, Queensland 4000, Australia

Abstract: This paper presents an integrated collaborative approach for facilities management (FM), which utilises Sydney Opera House as an exemplar case study. The project has three (3) research streams, *Digital Modelling, Services Procurement* and *Benchmarking*, which are considered individually and as an integrated whole. The project aims to achieve innovative FM strategies and models that will potentially be of benefit to the Australian facilities management industry.

The *Digital Modelling* theme aims to develop building information modelling for facilities management and investigates the potential of state-of-the-art information systems to enable a future integrated platform to support facility management collaborative activities and processes. The *Procurement* theme focuses on FM service procurement, especially for outsourcing. Procurement methods and a multi-criteria assessment approach for supporting decision making have been examined. The *Benchmarking* theme focuses on the maintenance of iconic assets, specifically performing art centres and facilities with similar functionalities. Critical success factors in the functional areas of asset maintenance are identified against organisational objectives of Sydney Opera House and key performance indicators developed.

Separately, the suggested collaboration between *Procurement* and *Benchmarking* includes potential for the linking of key performance indicators to support Sydney Opera House's Total Asset Management (TAM) strategies and corporate objectives.

Keywords: facilities management, digital modelling, procurement, benchmarking, total asset management

1 INTRODUCTION

A unique opportunity has arisen whereby Sydney Opera House, the Australian Government, the CRC-CI, FMA Australia and industry participants are in a position to jointly support Sydney Opera House as the FM Exemplar Project of the Facilities Management Action Agenda (2005). This project allows the facilities management (FM) industry to develop and showcase innovative methods to improve FM performance and promote best practice.

The FM Exemplar Project consists of three (3) key FM research themes: digital modelling, procurement and benchmarking. Whilst each of these themes represents a key focus area, the ultimate goal which distinguishes the project from previous research is the aim to link each of these themes together to produce an integrated Facilities Management framework utilising one of Australia's most recognisable icons.

This paper presents the background to the FM Action Agenda and FM Exemplar Project in Section 2 and provides the research status to date on Sydney Opera House. Specifically Section 3 deals with the development of Digital Modelling, Section 4 discusses the Services Procurement strategies based on the Sydney Opera House case study and Section 5 presents a benchmarking framework to establish benchmarks and identify the best practice for Sydney Opera House and the FM industry, which includes the development of key performance indicators and data collection. Section 6 provides insights into the collaboration between procurement and benchmarking through data sharing and references of key performance indicators and recommendations for an integrated FM solution.

2 DIGITAL FACILITY MODELLING

Sydney Opera House has a design life of 250 years and a very high quality of construction and finish appropriate for NSW's most prestigious entertainment facility. The building comprises 7 theatres, 37 plant rooms, 12 lifts, over 1000 rooms. Sydney Opera House employs 300 full-time staff and a further 300 part-time staff, delivering over 2500 performances per annum. Obviously to track and plan all the activities including facility management operations, such as regular and incidental maintenance, cleaning, building updates, etc., an information system is necessary. The digital modelling part of this project investigates the potential of building information modelling and state-of-the-art information systems to innovate and support facility management processes.

2.1 Sydney Opera House's Existing FM Systems

Commencing in 1958, building documentation was based on hardcopy paper and pencil/ink drawings, pre-dating even the introduction of two dimensional computer aided design (2D – CAD) technology. While excellent catalogues of this information are available, accurate data is extremely limited, in hardcopy, microfilm or digital format. At the end of Stage I of the construction in 1966, accurate surveys (in imperial units) of the current infrastructure of the ground works and podium was carried out and to date these are the definitive data of those parts of the building. The lack of consistent, reliable data has become a major problem after 30 years of occupancy, particularly many services system modifications, numerous small works projects and now, the need for significant refurbishment. Working without accurate drawings, Sydney Opera House has developed a pragmatic information management system, where the facility's characteristics are used as a tool. For example, many external pipes and ducts have been colour coded, objects such as equipments and installations have been labelled and almost all objects (doors, rooms, etc) have a unique ID. Several information systems for tracking maintenance operations and keeping track of assets have been installed and make use of the unique IDs.

The building has reached a milestone age in terms of the condition and maintainability of key public areas and service systems, functionality of spaces and longer term strategic management. The current "documentation" of the facility is comprised of several independent systems, some overlapping and is inadequate to service current and future services required. A number of key concerns are evident for this situation:

- The existing information systems are still "paper-based" which can result in outdated documents and data re-entering in different (independent) systems;
- the current information systems cannot be readily adapted to support benchmarking processes;
- the current information systems only handle limited procurement information;
- The information system comprises several distributed information sources containing a mix of current and redundant information;
- The information systems are not linked together. Linking or integrating the information systems will support the alignment of facility management operations with business processes. For example maintenance planning can be optimised by taking business activities into account.

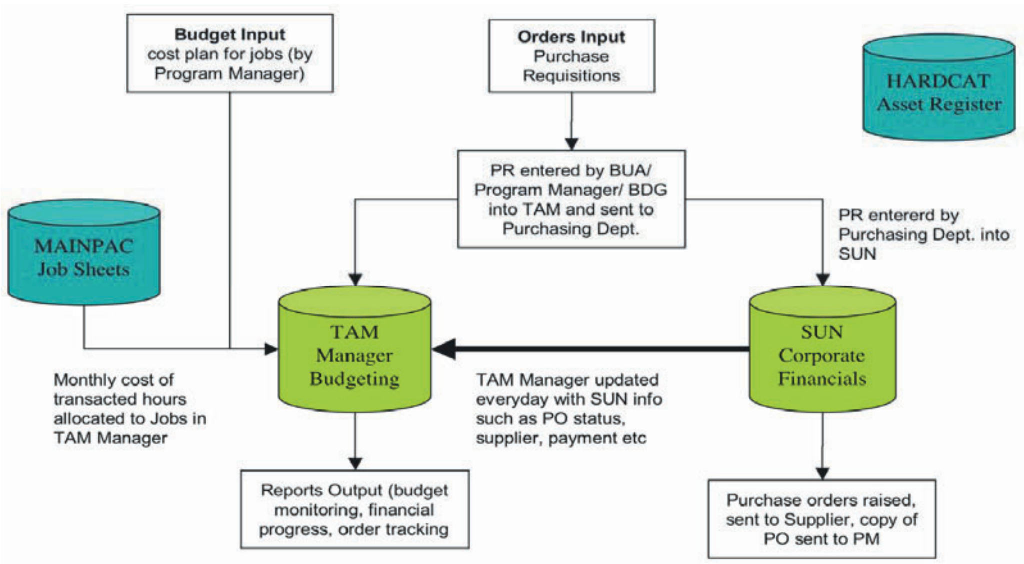


Figure 1

The key facility management systems currently used by Sydney Opera House.

2.2 Building Information Modelling (BIM)

An important consideration in this context is the use of an integrated model of the building, ie. a Building Information Model (BIM) to support in a comprehensive manner FM operations required by Sydney Opera House.

BIM is an integrated digital description of a building and its site comprising objects, described by use of an accurate three dimensional (3D) geometry, with attributes that define the detail description of the building elements, and relationships to other objects, e.g. a Duct *is-located-in* Storey Three of building Block B. The key generic features of BIM are presented as follows:

- *Robust geometry* - objects are described by faithful and accurate geometry, that is measurable;
- *Comprehensive and extensible object properties* - any object in the model has some pre-defined properties or extended properties to be customised by users. Objects thus can be richly described, eg. a manufacturers' product code, cost, or the date of last service;
- *Semantic richness* - the model provides various types of object relationships that can be accessed for analysis and simulation, e.g. is-contained-in, is-related-to, is-part-of, etc;
- *Integrated information* - the model holds all information in a single repository ensuring consistency, accuracy and accessibility of data;
- *Life cycle support* - the model supports the FM data over the complete facility life cycle from conception to demolition, extending current over-emphasis on design and construction phase.

The guidelines of BIM for Sydney Opera House have been developed stating how Contractors and other parties should supply their design information to Sydney Opera House. Adopting these guidelines enables Sydney Opera House to build up a digital repository of their facility which can be merged and re-used for different purposes.

3 SERVICES PROCUREMENT

The procurement research aims to review the current practices and make recommendations on an optimum procurement strategy for asset maintenance services and works. In particular, it will be developed from case studies of Sydney Opera House as an exemplar.

3.1 Procurement Strategies

Procurement strategies consider the achievement of strategic objectives, and operational requirements and constrains when deciding which procurement model is appropriate for providing a service. Figure 5 illustrates that the current Asset Maintenance Strategic Plan (AMSP) and the Operational Asset Maintenance Plan (OAMP) of Sydney Opera House generate 'demand statements' as evaluation criteria and strategic and operational requirements as performance criteria in the procurement process.

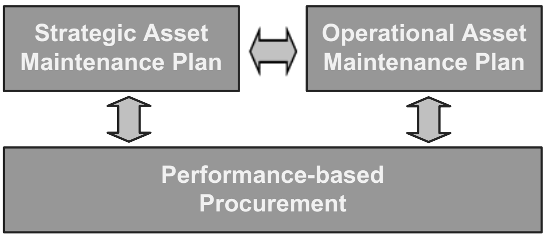


Figure 2

Illustration of the development of procurement strategies to meet the strategic and operational objectives of Sydney Opera House.

In the procurement process provision, options for required maintenance services and works are assessed and appropriate procurement routes are identified. The criteria of the multi-dimensional evaluation process consider the requirements formulated in the AMSP and the OAMP in the context of the prevailing market and the service provision reliability and risk, see Figure 6. The criteria are considered in the steps of the tender process by the various stakeholders of the process.

An important notion of this framework is the integration of a common evaluation system, founded on the strategic high-level objectives and the identified operational requirements, into the design phase and the sourcing, bidding and contracting phases.

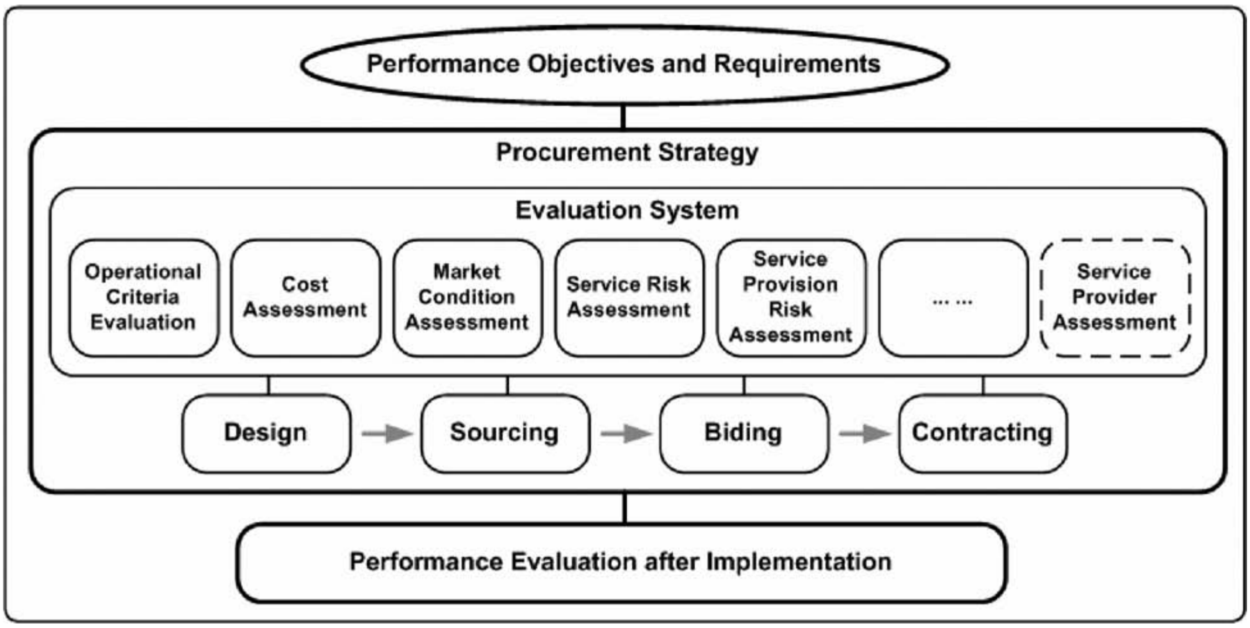


Figure 3

Development of a multi-dimensional assessment evaluation system for the procurement.

3.2 Procurement Vision

In order to ensure that the organisation's procurement approach is developed properly, it is necessary to outline how it will procure, i.e. to 'envision' targets against which performance could be judged. The following presents some examples of procurement visions:

- Maximise efficiencies and effectiveness in how Sydney Opera House makes purchases;
- Ensuring that procurement decisions are made on the basis of whole-life costs, cultural fit and not solely short term financial criteria;
- Ensuring that purchasing will be coordinated between departments where possible; in order to improve efficiency, appropriately planned and timed so as to increase overall value without increasing cost;
- Being fair and just in how and what it purchases and how the organisations 'treats' their Service Providers and Contractors;
- Providing leadership and building capability;
- Improving partnering and collaboration with our peers, and Service Providers and Contractors.

These visions can in turn be met by applying principles that provide strategic direction in how and what it procures, eg. minimising risks, being open and accountable, and being informed by customers' needs.

3.3 Procurement Principles

The following procurement principles may be developed to support the organisation's procurement strategies:

- Justification of business case;
- Developing clear and concise specifications, and good contract administration principles/guidelines;
- Appropriate balance between quality and price;
- Flexibility in developing alternative procurement and partnering arrangements, eg. fostering a culture of partnership, collaboration, and cooperation;
- Transparency and accountability;
- Compliance with legislation;
- Value for money;
- Equality of opportunity;
- Sustainability of procurement ('whole-life' consideration);
- Inculcate a culture of continuos improvement (ie. to provide for evaluation, improvement and change of circumstances during contract duration);

- Provision of professional work, moral and ethical attitudes.

3.4 Sydney Opera House Case Study

The current Sydney Opera House asset maintenance framework enables the following to occur:

- In-house (Facilities and Precinct Development Portfolio team): to focus on strategic issues, such as performance analysis, energy, environmental management and other strategies with the aim of resulting in better longer-term operational performance and reduced costs.
- Outsource (External Service Providers and Contractors): for issues other than strategic issues, with the aim of achieving long term efficiencies and cost benefits, e.g. maximised or increased life expectancy of plants, structures and major systems.

The procurement routes in Sydney Opera House for the following operational aspects are outsourced:

- Mechanical;
- Electrical;
- Building Maintenance;
- Fire Services;
- Security and Surveillance Systems;
- Transportation Systems;
- Stage Machinery Systems;
- Sanitary Plumbing and Plant;
- Consultant Services

A robust tender evaluation system supports an organisation making the appropriate assessment and selection of service providers and suppliers. Sydney Opera House has the following tender evaluation criteria for their cleaning services and select the appropriate criteria for each cleaning contract.

- Price;
- Compliance with specifications;
- Understanding requirement of work;
- General cleaning and management expertise (such as knowledge, communication skills, OHS, customer relations, standard of cleanliness, experiences, quality of references, capacity);
- Expertise in providing similar services;
- Quality of references;
- Capability to provide required services.

It is necessary for facility owners and managers to be able to continually monitor their progress and subsequently understand whether their Service Providers meet the user's requirements in terms of performance criteria and quality expectations. There should also exist the opportunity for continuous performance improvement and / or criteria modification over time to enhance satisfaction levels. Such information could be used by organisations to learn from their experiences and use them to facilitate their decision making process for future projects. Sydney Opera House sets out key performance indicators (KPIs) to regularly measure the performances of its service providers. Some examples from the contracts for the Maintenance of Mechanical Building Services are:

- No unauthorised disruption of operations;
- Work completed within time;
- Work completed to or better standards;
- Agreed budgets.

3.5 Findings

The trends in procurement appear to be pointing towards less adversarial procurement methods such as performance-based contracting, alliancing and relationship contracting which are believed to assist in the reduction of overall project costs and timeframes, to promote innovation and best practice, to enhance client / customer relations, user experiences and provide more satisfactory outcomes. However, all existing procurement methods may not be ineffective or under-performing. Services and works 'projects' are likely to possess a mix of characteristics from a range of contractual solutions. This affects the mode in which organisations may choose to select and manage their service provider and supplier relationships. There may be a need to examine multi-criteria decision-making tools/methods which enable organisations managing facilities to assess and select the appropriate providers in a more holistic manner.

Holistic assessment may be made with the aid of checklists where weighted scores are given to the criteria in relationship to their importance. The development of a multi-criteria analysis tool linked to data analysis referencing the KPIs could aid in the decision making process.

The diagram in Figure 4 addresses a relationship between an organisation's assessment of services procurement and benchmarking data collected, which could use KPIs as a link. The strategic decision determining the KPIs and the services and works procurement could be supported by the benchmarking data collected.

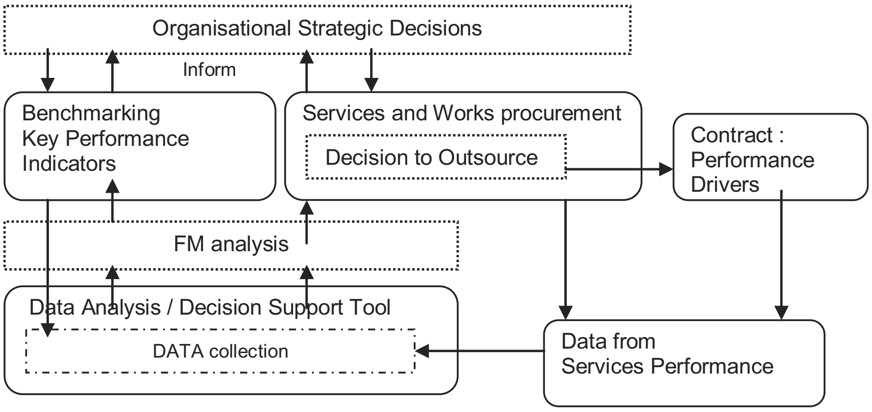


Figure 4

Illustration of the collaborative relationship between the procurement and the benchmarking data collected.

4 PERFORMANCE BENCHMARKING

4.1 Facilities Management Benchmarking

Clearly, the role of FM is to support the fundamental activities of an organisation or a specific facility(ies). Hence, the objective(s) of the FM functions should be compatible with and reflect those of the organisation.

Benchmarking systems in facility management can be employed to monitor, control and improve or to simply to rank the organisation and its assets according to its performance targets. Benchmarking can be performed by an organisation internally; or as an external exercise between comparable partners; or to assess and evaluate the development of a performance indicator over time. Figure 5 illustrates a proposed benchmarking system to assist Sydney Opera House in achieving their strategic objectives with respect to asset maintenance.

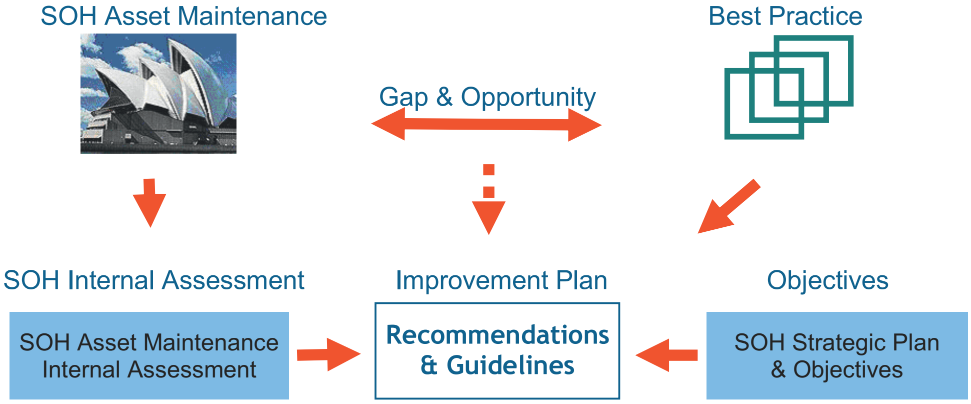


Figure 5

A proposed benchmarking system to assist Sydney Opera House in achieving their strategic objectives.

A classification scheme for applications of benchmarking in facility management was developed in this project, in order to be able to identify typical and especially successful benchmark types for specific objectives. The scheme is shown in Figure 6 for the example of the Building Fabric Index (BFI), as used in Sydney Opera House. The BFI is used with the objective of "monitoring and controlling": it is a quality benchmark, it looks at the trend of performance over time ("historic"), it is related to a specific function ("functional") and it has no driver ("none").

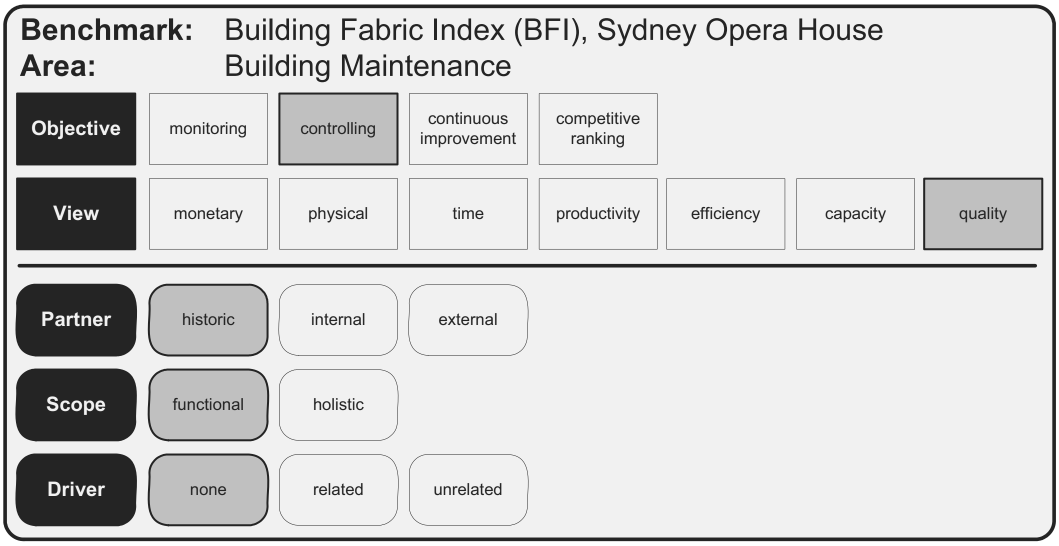


Figure 6

A classification scheme for benchmarks using the Building Fabric Index (BFI) of Sydney Opera House as an example.

4.2 Sydney Opera House Benchmarking Process

Based on the standard benchmarking process structure (Camp, 1989), a benchmarking process for Sydney Opera House cases has been developed, see Figure 7.

Figure 7

A benchmarking process structured for Sydney Opera House cases.

- *Identifying types of facilities and functionalities* refers to finding a group of facilities with similar characteristics, core business or comparable functionalities for benchmarking.
- *Choosing areas for benchmarking* selects features which are most valuable and relevant to Sydney Opera House for benchmarking.
- *Defining key performance indicators & measurements* describes performance objectives in terms of key performance indicators and establishes measurement methods and metrics.
- *Data collection and establishing common standards for comparison* refers to collecting data from a group of benchmarking partners and establishing common standards such as data structure and format for comparison.
- *Benchmarks and guidelines* identify benchmarks in the areas of interest and delivers recommendations and guidelines for Sydney Opera House and the FM industry in general.
- *Improvement plan* assists Sydney Opera House or the FM industry in improving performance in the asset maintenance area in terms of the benchmarking outcomes.

4.3 Development of Key Performance Indicators

Sydney Opera House has the primary function of a “Performing Arts Centre”. It is at the same time an “Architectural Masterpiece” and a “Heritage Building” and further is of “iconic” value for 20th century architecture and contributes to the tourism value of Sydney. These values bring objectives and requirements with them, which have to be integrated with or aligned to the objectives of the facilities management functions.

The development of key performance indicators (KPIs) is presented following a systematic structured approach as it is illustrated in Figure 8.

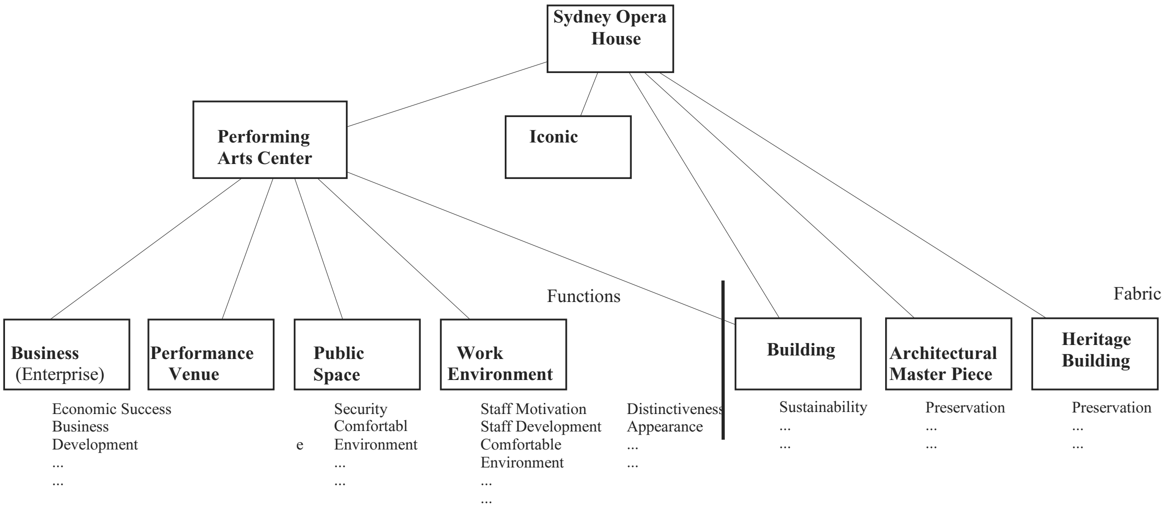


Figure 8.

Illustration of the KPI development.

Firstly the key organisational functions, and their corresponding function areas, their critical success factors and key performance indicators have been identified from available business documents such as reports and strategy plans. For instance, as a performing arts centre Sydney Opera House functions as a business with the objective of attracting and holding sponsors and partners. Its critical success factors include being attractive to sponsors and the branding of Sydney Opera House name while some key performance indicators may include the number and value of supporting sponsors. This is followed by the identification of the high-level objectives of facility management functions. According to the Sydney Opera House Trust Annual Report (2004), the key objective of facility management function is: “Providing first class venues, facilities and services that support our artistic and business aspirations.”

Secondly functions of facility management and its specific objectives are identified; for instance these include the building and custodial maintenance functions of facilities, are identified and their KPIs are derived.

Thirdly, these indicators are allocated to each of the perspectives. As an exemplar they have been categorist into four perspectives: financial; internal business processes; visitor and staff satisfaction; and innovation, growth and leaning.

The systematic objective development from high-level objectives to facilities management KPIs indicates which objective areas are relevant for the organisation and shall be employed to structure the benchmark system. Further it provides a hierarchical structured framework to identify the KPIs themselves.

4.4 Data Collection

The project made an initial contact with national and international organisations responsible for other “iconic” facilities seeking their participation with Sydney Opera House project as international comparators. Involvement will include an audit of the data that they are currently collecting, and it’s availability for benchmarking. The development of better data collection, quality assurance of data, standardised vocabulary, and comparable collection methods will be recommended to the participants.

5 TOWARDS FUTURE INTEGRATED FM SYSTEMS

The integration of different sources of information such as building information, maintenance information, service history and performance information, benchmark information, and business information can improve facility management operations in general. For example, queries retrieving bad performing objects can be cross checked with maintenance schedules. A scenario of a future integrated FM system based on the integration of information resources is illustrated in Figure 9.

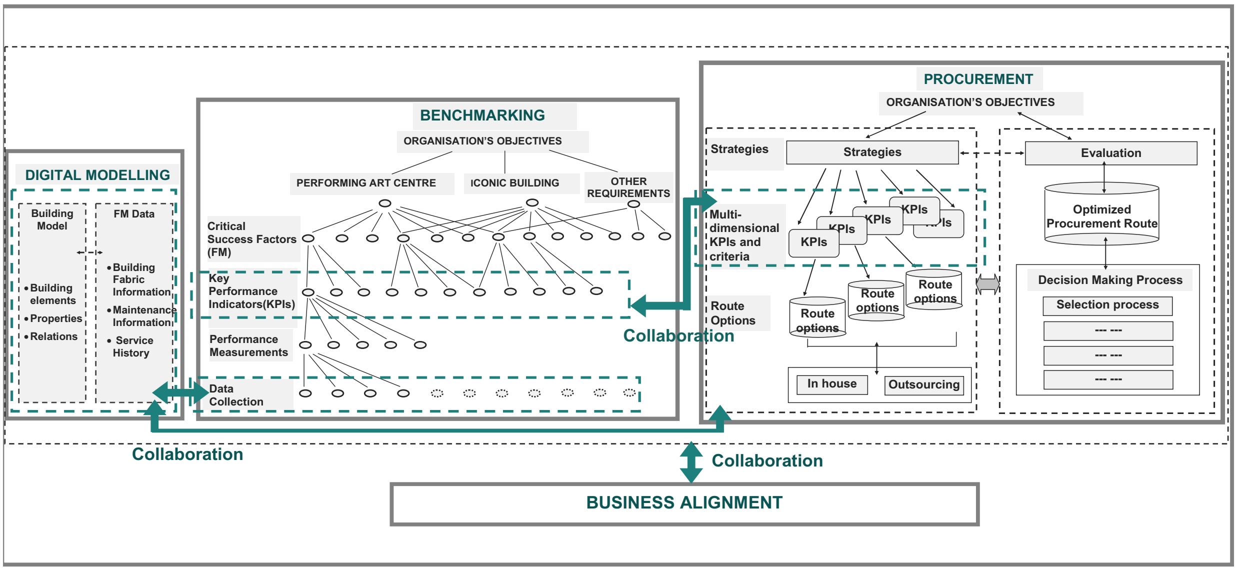


Figure 9

Illustration of a scenario of a future integrated FM system based on the integration of information resources.

It is suggested that the benchmark data collected be kept in a digital format for comparison and be able to integrate with the digital building models. Maintenance information, service history and performance, and business information are able to be easily retrieved from a data repository to enable an optimised maintenance schedule and procurement route. KPIs linked to the benchmark data repository can be utilised by Facility Managers to enable procurement strategies to be aligned with FM performance targets.

The future integrated FM solution should be supported by collaborative FM activities and processes. It should provide integrated information resources for sharing and use by Facility Managers, Contractors and General Management personnel. This enables improvement of collaborations on the operational level and faster and more effective FM processes, and executive reporting.

The future integrated FM system can be implemented using a centralised approach, or a decentralised approach. The centralised approach develops a central database that provides a common model for data sharing. The common model is complicate since it contains all relevant FM information to enable collaborative FM activities. The decentralised approach develops interoperability between different applications. Information sources can be communicated from one application to another for access and queries from different departments.

6 CONCLUSIONS

The Sydney Opera House FM Exemplar Project represents an excellent opportunity to leverage off the iconic nature of Sydney Opera House's international and national profile to identify and develop best practice within the FM industry. This project provides a broad range of practical input from client, consultants and service providers. The project's outcomes will in turn support the Australian Government's Facilities Management Action Agenda. The innovative methods delivered by this project should be implemented across the Facilities Management industry at the strategic, management and operational levels, with clear training and educational benefits leading to improved service delivery for the benefit of the community.

7 REFERENCES

[1] Brackertz, N. and Kenley, R. (2002). Service Delivery Approach to Measuring Facility Performance, *Facilities*, vol.20, no.3/4, pp.127-35.

[2] Camp, R.C. (1989) *Benchmarking: The Search for Industry Best Practice that Lead to Superior Performance*, ASQC Quality Press, Milwaukee, Wisconsin, U.S.A..

[3] Facilities Management Action Agenda (2005), *Managing the Built Environment, Commonwealth of Australia*, Australian Government, Department of Industry, Tourism and Resources, Canberra.

[4] IFC 2x Extension Modelling Guide, http://www.iai-international.org/.

[5] Kaplan, R.S., Norton, D.P. (1992) The BSC: Measures that drive performance, Harvard Business Review, vol.70, pp.71-9.

[6] Sydney Opera House (2005) Sydney Opera House Building Presentation Index (BPI) and Building Fabric Index (BFI) - Functional Specification for Software Development, Internal Document, Sydney Opera House.

[7] Sydney Opera House (2004) Building Maintenance Contract, Internal Document, Sydney Opera House.

[8] Sydney Opera House (1999) Cleaning Services, Internal Document/Contract, Sydney Opera House.

[9] Sydney Opera House (2003) Conservation Management Plan prepared by James Semple Kerr, Internal Document, Sydney Opera House.

[10] Sydney Opera House (2002) Strategic Asset Maintenance Plan (period 2003/04-2027/28), Internal Document, Sydney Opera House.

[11] Sydney Opera House (2005) Total Asset Management Plan (period 2006-2031) prepared by Rider Hunt Terotech, Internal Document, Sydney Opera House.

[12] Sydney Opera House (1998) Tender document for mechanical building services Maintenance, Internal Document, Sydney Opera House.

[13] Sydney Opera House Trust (2004), 30 years of inspiration, Sydney Opera House Trust Annual Report 2004, Sydney Opera House.

Acknowledgements

In addition to the contributions of the Project Team, the author wishes to acknowledge the support of the FM Action Agenda Implementation Board and the Facility Management Association of Australia, for their valuable advice and assistance on identifying national and international benchmarking participants, Ron Wakefield of RMIT, Nicholas Ferrara of Rider Hunt Terotech, David Marchant of Woods Bagot, Andrew Frowd of Queensland University of Technology, Selwyn Clark and Frank Seed of Queensland Department of Public Works for their participation and assistance on this project.

Project Team

Cooperative Research Centre for Construction Innovation [1]

CSIRO Manufacturing & Infrastructure Technology, Australia [2]

University of Sydney, Australia [3]

Sydney Opera House, Australia [4]

Transfield Services, Australia [5]

Rider Hunt Terotech, Australia [6]

Facility Management Association of Australia [7]

CRC-CI [1:]Peter Scuderi, Kylie Legge and Marcello Tonelli (March to May 2006)

CSIRO [2:]Lan Ding, Robin Drogemuller, John Mitchell and Hans Schevers

USYD [3:]David Leifer, Dirk Schwede, Janet Henriksen and Jeremy Wu (until April 2006)

SOH [4:]Paul Akhurst

TSA [5:]George Spink

RHT [6:]Stephen Ballesty, Jason Morris (April – Nov. 2005) and Ankit Shah (from May 2006)

FMA Australia [7]

PROFIT CONTRIBUTION MAPPING

Mike Sondalini

Lifetime Reliability Solutions, 2 Denise Close, Shelley, Western Australia, 6148, Australia

Abstract: *Profit Contribution Mapping.* This paper introduces an enhanced financial technique to proactively identify and display operating costs and profits so operations and senior management can make decisions that deliver the best use of plant and equipment assets. Maximising profit from a business requires both efficiency and effectiveness from the production operation. An effective production process makes and delivers what the customer wants. An efficient production process delivers the profit the business shareholders want. Profit Contribution Mapping extends the value stream mapping concept and introduces financial measures to identify where profits are made during production and where they are lost. It combines accountancy, economics, engineering, maintenance, operations management and asset management to create a powerful management tool for effective control at the 'global' production process level and also efficient control at the 'local' process step level.

Keywords: value stream mapping, profit contribution mapping,

1 INTRODUCTION

A common difficulty encountered by engineering asset managers, maintenance managers and reliability managers is to convince 'bottom-line' focused Chief Executive Officers, senior managers and operations manager of the great value in adopting proactive reliability-based equipment care, use and purchasing practices. Because such practices can take years to show a financial return they are not easy to introduce into organisations with close horizon financial focus. Often financial management people simply claim that there is no clear economic sense in introducing such practices because the net present value of an initiative does not justify the capital expenditure required. Production managers can quickly claim that their current practices are suitable and it only needs people to 'sharpen their act' and all will be fine. Little do they realise the tremendous financial waste that such thinking causes their organisation now and in the future.

Because there are no suitable predictive financial or accounting tools available to readily explain to the non-asset management-aware person of the tremendous cost and waste resulting from poor plant and equipment operating and management practices, it was necessary to develop such a fnancial tool. The financial tool explained in this paper, Profit Contribution Mapping, shows an organisation the detailed consequential financial costs of using and persisting with current organisational and production thinking, practices and habits. It allows managers to perform 'what-if' scenarios by modelling the effects of different operating and asset management decisions to let them optimise results for the operation.

The value stream mapping method has proven to be an accepted method to influence production and senior managers on the worth of eliminating waste from their production processes. Because of the solid acceptance this technique has in industry it was selected as the base on which to develop a predictive financial model for justifying the adoption of sound facility management and world-class production practices.

2 PRODUCTION PROCESSES

A production or manufacturing process should be viewed as a system for supplying the customer's requirements. A system implies many interrelated parts working seamlessly together in a coordinated effort to provide an important and necessary function.

A production system involves the seamlessly operation of people, equipment and practices working in an ordered way to produce what the customer wants at a sustainable profit for the producer. An important job of managers, economists, accountants and engineers is to develop business systems that reliably and ethically achieve this seamless operation to the benefit of the organisation, its customers and the community. It is desirable to continually improve and tune the organisation to make it more efficient and do its functions faster, better and cheaper.

Production, processing and manufacturing systems turn raw materials into finished products through a series of steps that progressively convert them into saleable products. The conversion process typically requires added inputs such as labour, utilities like power and water, specialist services like maintenance, added materials like boxes for packaging, along with other numerous requirements to make products customers buy.

Production process and indeed service operations can be symbolized by a block diagram containing a series of boxes for each conversion step in the process, with materials, utilities, services and labour shown by arrows to represent their flow. Such a symbolic production, manufacturing or service process used to convert raw materials to customer-desired products is shown in the process flow diagram of Figure 1.

In Figure 1 raw materials are the direct materials used in the production process. The added inputs include the utilities (power and water) and services (such as boxes for the product, labour man-hours, lubricants for machines, etc) needed to complete the process step. The process steps use these to add value and make the products produced by the organisation. During production the product increases in value equal to the value added in each conversion step. Each value-adding step contributes part of the profit made when the product is sold to the customer. A process step does not produce perfect conversion and some losses occur.

From Figure 1 a few simple properties of a production process can be observed:

i. The process design establishes the process' capability to make the product.
ii. Product quality is determined by the process design.
iii. The bottleneck limits the maximum throughput rate for the process.
iv. The efficiency of operating each process step determines it profit contribution.
v. The customer demand rate dictates the product manufacturing rate.

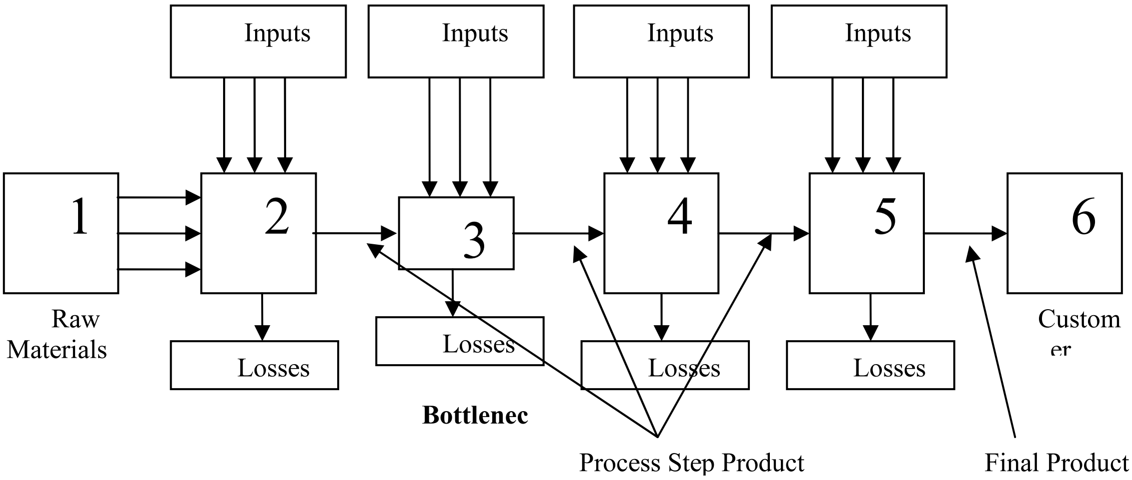


Figure 1. A Production Conversion Process

3 PROPERTIES OF PRODUCTION PROCESS

A production process should only make what the market will purchase. A production system has inbuilt natural checks and balances that keep a production process in-step with the market for its products. The rate of sale of the product to the customer is naturally controlled by what the customer wants and when they buy. It is important that production does not to make more product than can be sold; otherwise money is tied-up in product and inventory that no one wants.

The amounts of raw materials put into a process are governed by the market demand, the production usage rate and the ability of the business to pay for them. Similarly the added inputs throughout the process are governed by the individual process step needs and the cash flows available to pay for them. Where the customer demand rate is above the process' capacity to make the product, they are brought at the rate that the process can use them and can afford to pay. Market economics and business economics act to regulate and control the production rate.

This is the essence of a market-based, capitalist economy – the manufacture of products that people want in production systems that are balanced to the demand.

There is one more issue to consider in a production process - the losses. The process losses behave differently, very differently, to anything else in the production process! They are not naturally limited by demand. They are only limited by how much money is available to the production system. Because there are no systematic internal constraints on waste they are actively managed by minimising them at the design phase of the process and by managing them to minimal levels during the long operating phase of the business.

4 BOTTOMLESS PIT OF LOSSES AND WASTE

Usually not considered seriously important in business process design are the amounts of losses produced – the waste! The wastes include the obvious waste product and scrap materials commonly associated with production waste. But there are many other types of waste produced. All wastes take money from the profits that could have been made.

The other wastes, which are numerous and common but not often noticed, include such things as excess movement, lost heat, lost water, lost energy, excess storage space, excess in-process inventory, excess time, lost time, quality defects, excess forklift pallet hire, excess equipment hire, safety incidents, environmental incidents, excess paperwork, excess manning, and many, many more. (See Appendix 1 for a list of common costs in a production process.) Figure 2 shows the same process as before and includes all the wastes from the business and production process.

Some of these wastes are identifiable by using value stream mapping analysis. Profit Contribution Stream Mapping spots all wastes and gives them the financial value lost to the business to justify and encourage their rapid elimination.

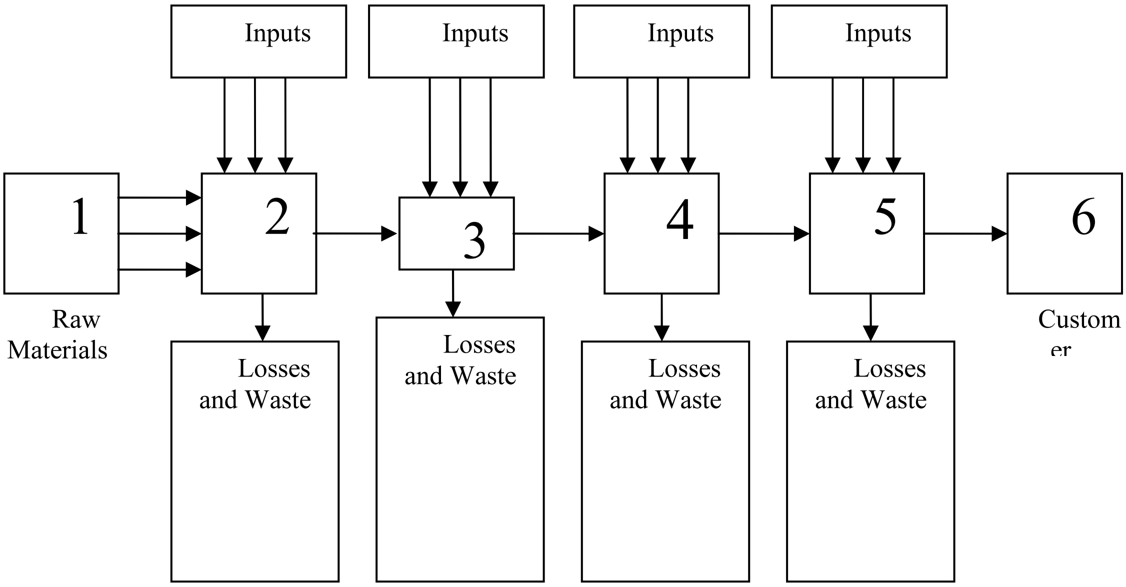


Figure 2. Losses and Wastes in a Production Process

The only natural means of waste control for a production system is how much money can be spent on raw materials and added inputs. Since creating waste has no natural means of self-control beyond bankrupting the business, it becomes necessary to develop business control systems that monitor the waste and force its minimisation and eventual total elimination.

There are now two other properties of a process that was not clearly evident before:

vi. Wastes extract effort and profits from a process.
vii. The process can turn raw materials and inputs into waste so that the process makes waste instead of profit, to the point where waste consumes all the profits.

Based on these seven properties, a business' production process can be interrogated to understand how money behaves within it and to identify the wastes that reduce its performance and profit. This analysis process is called Profit Contribution Steam Mapping (PCSM).

Once a process is operating, people's concerns naturally turn to making the product on time. What most people forget to do is to also make it efficiently, while their meeting customer's requirements. The demand to make product on-time often overrides the need to make it cost effectively. This leads to the situation in business where everyone is busy making product, but no one is busy making profit! If this situation occurs in an organisation, the creation of waste instead of profit dramatically rises. PCSM is a new way to help manager and engineers collect the cost information needed to operate a production system efficiently and effectively.

5 IDENTIFYING THE PROFIT CONTRIBUTION STREAMS

Figure 1 indicated that there are discrete steps in make a product and delivering a service. Each step has its own raw materials, which is the feed from the prior process step. It has its own added inputs needed to make the conversion. From each step come a ‘product’ and the wastes. Each process step is clearly identifiable from its predecessor and its successor and is self-contained in performing its conversion.

Since each process step is independent of the others it can be taken in isolation and viewed as a whole system in itself. This allows analysis of the process step separately. To make it clear which process step is being reviewed you draw a boundary around it on the process flow diagram. An example of segregating a process step for analysis is shown in Figure 3.

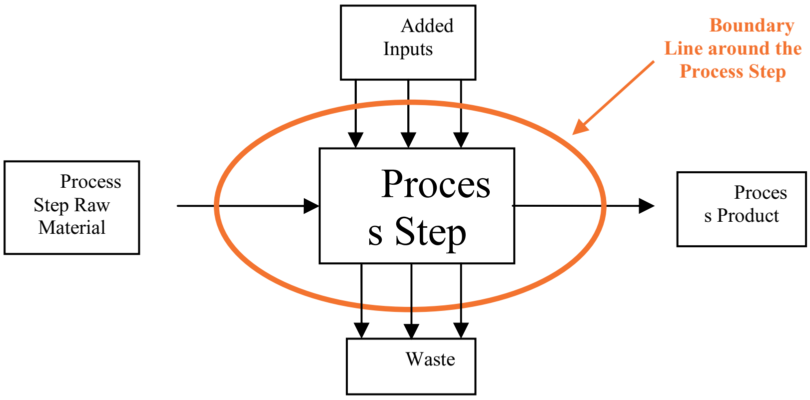


Figure 3. Local Process Step Profit Contribution Map

To determine process effectiveness and efficiency we need a measure. A good measure to use is money. Money is the universal language of business and most people understand the concept of using money to value an item or service. By using money to measure process product cost, added input cost, cost of wastes and the profit contribution from a process step, we can trend the process profit contribution step by step as the product is made. Figure 4 indicates the various money flows, both in and out of a production process.

Though only the money flows for Process Step 2 are shown, they equally apply for each step. By analysing each process step, the true costs of the raw materials, the additional inputs, the wastes from it and the total process step profit contribution can be determined.

Monitoring the costs and profit contributions of each process step provides a means to measure the efficiency of the conversion processes. The more profit contribution generated from a conversion step the more financially efficient is the step. By knowing the cost of all inputs and all wastes, the profit contribution made to the total profit by each process step can be identified. With profit contribution information managers, accountants and engineers can focus on new cost reduction, productivity and process improvements that return the best value for the operation.

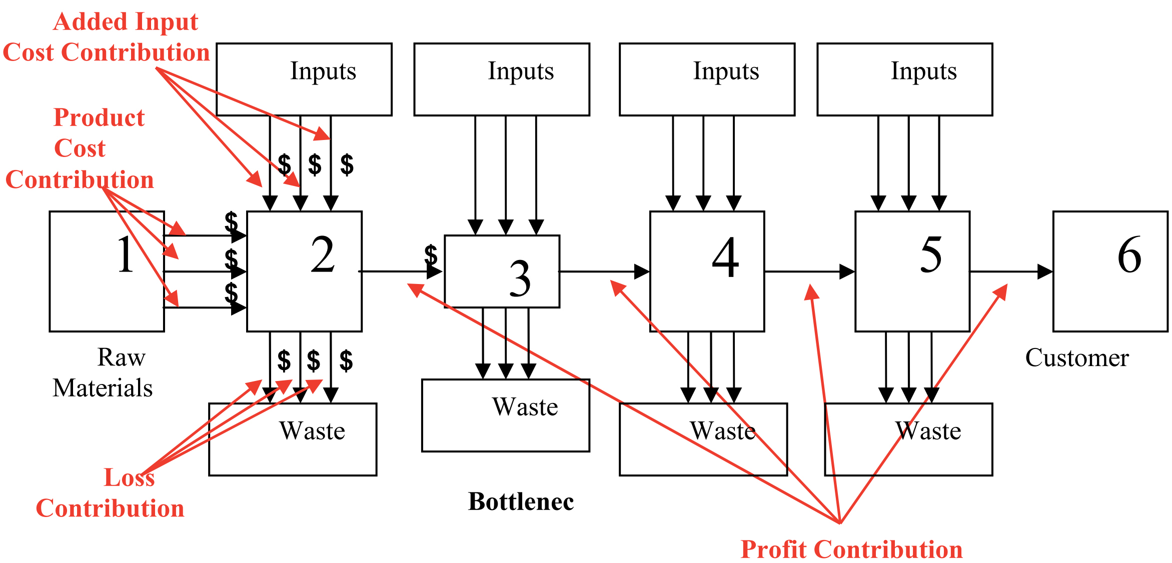


Figure 4. Process Profit Contribution Mapping

Figure 5 indicates how to identify each money flow associated with a process step. The boundary line makes it clear there is money coming into it from ‘raw materials’ and the added inputs required to make the process conversion. Each process step delivers its own process ‘product’ with its profit contribution from the value-adding performed in the step. In addition there are lost moneys that reflect process and operating inefficiencies, wastes and losses.

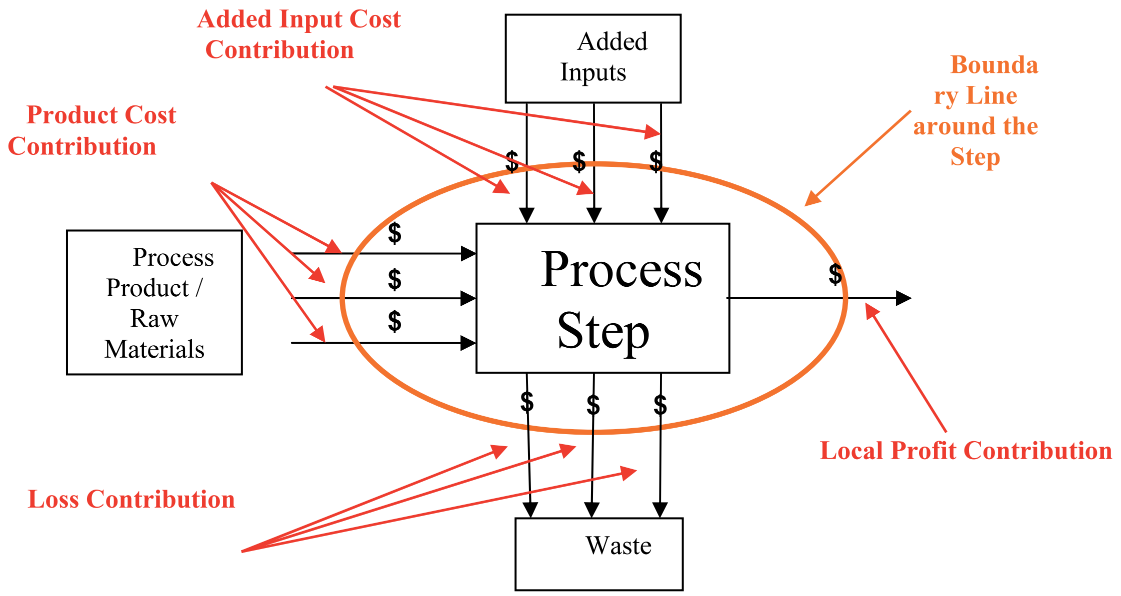


Figure 5. Local Process Step Profit Contribution Map

Figure 5 makes clear the importance of cost control in each process conversion step. If so much money is wasted in each step that the customers will not pay what it costs to make the product, there will be no sales and the organisation goes out of business. For products of equal quality it is very likely that the customer will find and use the cheaper producer.

With all the in and out money 'flows' in a process step identified they are used to analyse the profitability of the process step by calculating the costs and losses to determine its profit contribution. This can be represented by a simple equation.

Cost of Process + Cost of Added Inputs = Local Profit Contribution + Cost of Waste *(1)*

Product Input

Alternately the equation can be written as:

Cost of Process + Cost of Added Inputs - Cost of Waste = Local Profit Contribution *(2)*

Product Input

Strangely, from equations 1 and 2, it could be said we pay for waste twice, once when we buy it as an input and second when we throw it away as lost profit.

5.1 COST ANALYSIS

The power of Profit Contribution Stream Mapping is the clear understanding of production money flows it delivers through the use of visual management. A display of the business process is developed to indicate where the money goes into, around and out of the business. What happens with costs and profits is clear and issues highlighted by the mapping can be addressed.

5.2 ACTIVITY BASED COSTING

Activity Based Costing (ABC) is the most appropriate accounting technique to apply when determining process step costs. Standard costing is not suitable since overheads are allocated to direct costs. Rather it is necessary to capture every cost, from the smallest cost to the largest, as it is really spent. A total and true reflection of what happens in each process step during a set period of time is necessary.

Because the profit stream costing process identifies every cost individually, overheads allocation is not permitted. All overheads are identified separately and costed in exact proportion to their usage in each process step. If overhead costs were allocated to direct labour an incomplete mapping of the true costs would result.

The accuracy and completeness with which the process step costs are collected will directly determine the effectiveness of the profit contribution map as a management control tool. If data is complete and true then good, reliable business-improving decisions will be made to make the operation more efficient and profitable.

The appropriate time period to be used to collect the mapped costs may be an hour, day, week, month and even a year. The time period is dependant on the cycle time for the process step and the size of the time window necessary to identify all money flows for the process step.

Applying ABC permits identification of every cost with its component costs, and even sub-component costs. It is important that every dollar spent in the production of goods is accounted for and shown on the profit stream contribution map.

5.3 COLLECTING COST DATA

The cost data needed to analyse and manage a process is typically generated by the process as it produces its product. The cost of materials, labour, utilities, overheads and services are found on invoices or payslips. What is not normally available are the process costs accurately allocated to the process steps that incurred them. To manage a process step's efficiency it is necessary to cost every one of its inputs, products and wastes accurately.

An approach used to identify the money flows in a process step is to take the process step procedure and work through it. As it is read the process step raw materials, the added inputs, the wastes and the produced product are identified.

As shown in Figure 3, a boundary is drawn around the process step to clarify its associated 'flows'. Many of the inputs, wastes and products are shown on the process design drawings, or found in engineering documents, equipment manuals and in standard operating procedures. The data is confirmed by also personally observing the process step for a full cycle of production.

When onsite identify all electrical power supplies to the equipment, all pipes suppling services, all process products into the step, all added inputs into the step, all outputs and wastes from the step. This includes measuring the manpower, management, supervisory and maintenance efforts, times and costs incurred by the process step. It includes measuring forklift movements, vehicle movements, personnel movements, etc. that occur in the time period observed. It includes counting the number of lights and time they are on, how often equipment is hosed down and the amount of water used. All activities are collated and costed in a spreadsheet.

It will be necessary to go as far as identifying minor costs, like rags used for cleaning equipment and the cleaning detergents used. Another example would be to identify the use of personal safety equipment and company brought clothing each operator requires during the time period. Over a year these minor expenses can grow into serious costs that are easily wasted.

Find every dollar that goes into a process step and that comes out of it. Put on the mantle of the television movie crime investigator and look for all the clues to the puzzle. Unearth the truth, the whole and total truth of where money goes in each process step.

When studying a process step that involves movement of product and/or people, for example storing materials in a warehouse, time the length of the move, measure the distance moved and identify the equipment used in the work. Put a cost to the movement of product and materials so that it can be tested to see if it delivers real value for the expenditure.

All the process costs can be found in the business systems such as payroll, inventory and accounting. Unfortunately they most likely will be totalised costs. The labour will be for a person's total time at work and what is needed is what they spent in the process step. The power bill will likely be for the whole of a building, whereas the cost of lights and power for a machine in that building is required. The purchase of safety gloves will be in batches of dozens at a time but it is necessary to know how many are used by the people working in the process step.

The most accurate approach is to get the real usage of inputs and wastes. For example, the power used by the lights and machinery in the process steps need to be collected for the time period concerned. If that is not possible it becomes necessary to proportion the machine's share of the building's power based on the electric wattage stated by the manufacturers of the equipment used in the process step. By proportioning inaccuracies will be introduced that may eventually cause people to question the final conclusions.

If necessary introduce special means to capture cost information. Time sheets and record-of-use sheets can be developed, chart recorders can be connected to electrical equipment and meters can be installed to measure flows in pipes. When accurate cost control is important to the success of a business spare no effort to discover the total true costs of production.

5.4 LABOUR

- ***Direct Labour*** comes from the time sheets of the people employed directly in the process step being analysed. If the people are used in another process step then only cost time expended in the process step being investigated. The direct labour cost is the pay rate paid to the people working in the process step, multiplied by the time they spend in the process step during the selected time period costs are being collected for. Include all their on-costs such as allowances, superannuation, benefits, etc, proportioned to the period. Do not include allowance for overheads, as they will be separately identified.

- ***Indirect labour*** costs are the time spent by persons, other than the directly involved people, whose services are needed to ensure the process step is completed. It is necessary to measure and allocate times for indirect labour. This includes such costs as maintenance, supervision, middle and senior management time, inventory and storage personnel, purchasing department personnel, quality control personnel, etc. Identify these costs by interviewing relevant people to determine the time they spend in the process step. During a site inspection watch the process for a full production cycle and observe who interacts with the process step.

The indirect labour cost is the pay rate paid to the indirect people, multiplied by the time they spend in the process step during the selected time period. Include all their on-costs such as allowances, superannuation, benefits, etc, proportioned to the period.

If a short time period is analysed, say a week, and not all indirect labour is captured, it is still necessary to allocate a proportion of all the indirect labour costs to the time period. In this case take a longer time period, say a month or quarter year, and collect all the costs for the longer period then proportion and allocate them in weekly quantities.

- ***Indirect expenses*** are those costs incurred due to the presence of the people in the operation. An example is a manager's car and fuel which is paid out of operating revenue. Allocate them in proportion to the hours spent in the process step by the expense owner.

5.5 SUBCONTRACTORS

Subcontract labour costs and materials used in the process step need to be recognised and allocated in the same way as employed direct labour. There will be an invoice for the subcontractor's time and materials and the allocation of times and materials for the work done in the process step can be extracted from it.

5.6 UTILITY SERVICES

Electricity, water and such services will need to be measured and allocated to the process step usage during the time period.

5.7 MANAGEMENT, ENGINEERING, ADMINISTRATION AND SUPERVISORY COSTS

These costs cover the time managers, engineers, supervisors and administrative support staff spend doing work related to requirements of the process step. For example daily meetings, site inspections, human resources requirements, problem solving process issues, invoicing matching purchases, maintenance planning, etc. All support persons who interact with the process step need their times and costs recorded against the step.

Initially interviews can be held with people to ask them to estimate the time they spend on the process step. If necessary have them keep time sheets to record the actual times spent involved with the process during the time period.

5.8 ADDED INPUT MATERIALS

Direct material costs are for added input materials actually used in the process step. They are the obvious additions of substances into the process step. This includes such things as electricity for motors, boxes for packaging, lubricant for equipment gearboxes, air for pneumatic rams, etc. Typically these materials can be identified as entering the process step in a physical form. The parts of them wasted can be measured and given a value.

These costs depend on the quantity and value of each input material used during the time period. It requires counting the amount of the material used and multiplying by the cost of the added material. Material costs can usually be identified from invoices for the material. Sometimes the added material has been made by the organisation and no invoices are available. In such cases it will be necessary to get an accurate cost for it from the organisation. If none is available it will need to be calculated from the cost of the labour, ingredients, handling and manufacturing charges used to make it.

Indirect material costs are the costs associated with the indirect functions required to perform the process step. Such as paper for recordkeeping, electricity for lighting, a maintenance planner's computer, the cost of forklift hire to move pallets, the building storage space for spare equipment parts, etc. All these costs are real costs incurred to conduct business that supports the production processes and need to be identified.

It is necessary to measure them and quantify them so that they can be given a value. Measurement can be by stopwatch, distance, counters, etc. Their use in a process step needs to be identified and quantified to reflect how much is used in the process step conversion and how much is wasted.

5.9 PRODUCT COSTS

The cost of a product entering a process step is needed. An accurate value may be available from the accounting, or production department. If it is not available accurately it will need to be calculated for each prior process step commencing with the start of production.

5.10 IDENTIFYING AND COSTING WASTES

Direct waste is any unused direct labour or direct materials added into the process which are not fully used in making a product. Even if the added input is gradually converted through a number of process steps, as long as it is fully used it is not waste. Unconverted added input is waste.

For example, in some chemical processes the chemical reaction absorbs only a portion of the mixed ingredients. Those ingredients that are not converted by the reaction are wasted. A laboratory can analyse for the unconverted ingredients and tell how much was unused.

Another example is water used to clean equipment. If the water is not fully used in the process to make product but disappears out of the process, then it is wasted. Leakage out of the process is waste. Spillage from a process is waste. Another example of waste is side-steam materials collected in bags or bins to be disposed of outside of the process.

Indirect wastes are those wastes that relate to the unnecessary use of indirect labour and indirect materials. They are more difficult to identify because they are not easily observable. Examples include wastes related to lost time in meetings, to lost energy, to lost compressed air, to safety equipment thrown away before being fully used, to storing unneeded materials in a storeroom. There are numerous instances of such wastes.

The detection of indirect wastes is through observation. That is why it is necessary to be present during a full cycle of a production process and observe all process steps and their inputs to identify wasted costs, materials and product. Look in the rubbish bins used in the process step area of the business and see what is thrown away. Are lights and air conditioning left on overnight unnecessarily? If required develop and instigate systematic means to spot and record the waste and its value during the period investigated.

5.11 COMPARISON WITH STANDARD COSTS

Every organisation should have a standard costing system for its products. If there are standard costs available compare them with the costs from the profit contribution mapping analysis. Using existing standard costs double-checks the analysis and concerns can be raised when costs are far from the standard costs previously allocated.

5.12 USE OF COMPUTERISATION AND TECHNOLOGY TO CAPTURE COSTS

The work involved in identifying and costing component inputs, products and wastes for each process step can be large. If a company choses, it can use modern technology and computerisation to capture many of the costs automatically. Labour can be identified electronically by using electronic cards and time clocks. Materials can be identified electronically via electronic tagging or bar coding.

Wastes are difficult to identify electronically. It may be necessary to change work procedures and include the recording of waste. Introducing counting and measuring of wastes will allow people to identify the causes and address them before they get even worse.

Through the use of Global Positioning Systems equipment, materials and people movements can be tracked and any time losses identified and addressed.

5.13 MANAGING WITH PROFIT CONTRIBUTION MAPPING

Once all the flows into and out of a process and its steps are identified and costed the profit contribution map can be created.

6 THE PROFIT CONTRIBUTION STREAM MAP

A simplified partial sample of a Profit Contribution Stream Map is shown in Figure 6 for a section of canning line. All the costs are shown cascading to and from the process steps. The time between updating the map depends on the urgency of the situation. If it is necessary to make changes quickly, then more updates are needed than if the situation is stable.

At least every month the profit contribution map ought to be updated. If the mapping process can be done electronically it would be useful to do an update weekly, as this frequency allows better control of the operation. In some cases it might be necessary to map a particular process step more often than the entire process because of the importance that step has in the operation.

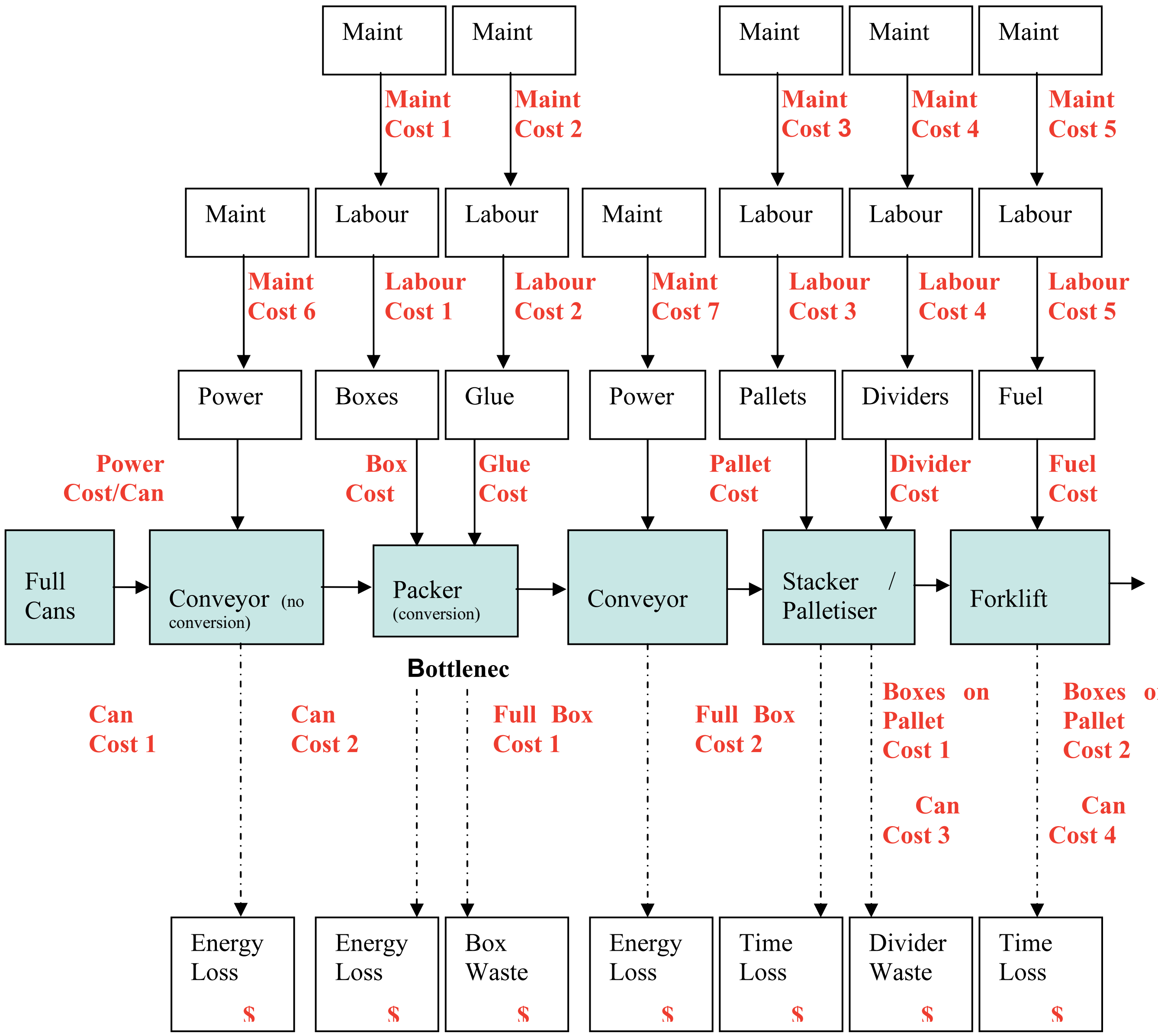


Figure 6. Profit Contribution Map for a Canning Line

6.1 PERFORMANCE MEASURES AND REPORTING

The problems that are highlighted by the profit contribution analysis can be addressed by the business management and personnel with new strategies applied to maximise the profit contribution. After a process step is analysed in detail it is easy to understand and appreciate how its many factors interact and impact each other. The accurate costing of inputs, wastes and conversions will identify problems. Through detailed questioning the reasons can be uncovered and the required changes made.

If change is required it is necessary determine what that change will be. The issues will need to be discussed with everyone concerned to fully appreciate and understand the situation’s history and reasons for occurring. The new changes will also need discussion, review and analysis for possible unwanted consequences. Finally a decision will be taken and changes will be made. When the changes are introduced they too will need to be measured, monitored and reported.

6.2 PERFORMANCE MEASURES

Selecting the right measures to monitor and report will be critical to the success of the change process and to the speed of its implementation. The measures need to be meaningful to the users, truly reflect the situation, are within their control to improve and inspire continued improvement.

One of the change strategies will be to introduce performance measures that identify poor efficiencies and the practices that cause them. Performance measures based on the issues identified by the analysis will drive the right behaviours and actions from people. The measures can be graphed and trended to show performance improvement.

Some typical measures to use are listed below. Measures can be developed that are suited to specific circumstances. The purpose of measuring is to know exactly what is happening. Once the current situation is clearly understood an assessment can be made as to whether it is satisfactory or it needs to be changed. If a change is made the effects will be seen in the performance measures. It may take several months for the effect of a change to be observed. Where the measures indicate an unsatisfactory result, a correction is necessary to get on-track.

- ***Usage Efficiency:*** This is the classic output divided by input. Select the process flows that are important and develop an appropriate efficiency measure for each and trend them over time.
- ***Productivity:*** These are measures of process performance. They are time based ratios of output during the time period. From the profit contribution map select the productivities which are important to measure. Productivity can be measured at the process step level and at the global process level.
- ***Throughput:*** This measure is a count of what passes a selected point in the production process during a period of time.
- ***Waste Cost:*** This measure counts the cost of waste in dollars per dollar spent to purchase the original material.
- ***Quality:*** This is the proportion of production that meets customer specification. It is another measure of a wasteful process.

To get a complete understanding of what happens in a process requires more than one measure. Business processes involve many interactions that affect each other and it requires a number of ratios to identify what is occurring in a process. Do not use any more measures than necessary. Maintaining measures requires time and money, which are then not available for use elsewhere. Experiment with the right measures to apply before deciding which ones to keep and use.

6.3 PROFIT PERFORMANCE REPORTS

Keep reports simple by using headings to categorise the report and visual means for displaying information. Show trends graphically in a suitable form to make their message clear to users. Use balloon notations in the graphs to highlight issues that need attention. Apply colour and font variations to enliven the report.

When a table is required to list details show summary entries and totals for each category. Keep the details for when people ask. Draw people's attention to the conclusions and their implications by providing an executive summary at the top of the report.

7 CONCLUSION

Engineering asset management is as much about the wise use of money as it is about the wise use of operational management, engineering and maintenance to deliver top performance from production equipment and processes. By fully understanding the costs and money flows in a production operation, its managers and employees can make good decisions that are effective and efficient for the well-being of the business, the shareholders, themselves and their community.

8 APPENDIX

8.1 THE MANY COMPONENTS OF PRODUCTION COSTS

- Labour : both direct and indirect
  - operators
  - repairers
  - supervisory
  - management
  - engineering
  - Overtime / penalty rates
- Product waste
  - scrap
  - replacement production
  - clean-up
  - reprocessing
  - lost production
  - lost spot sales
  - off-site storage
- Services
  - emergency hire
  - sub-contractors
  - travelling
  - consultants
  - utility repairs
  - temporary accommodation
- Materials
  - replacement parts
  - fabricated parts
    - materials
    - welding consumables
    - workshop hire
  - shipping, loading, transport
  - storage
    - space
    - handling
  - disposal
  - design changes
  - inventory replenishment
  - quality control
- Equipment
  - energy waste
  - start-up
  - shutdown
  - handover/hand back
  - decommissioning/commissioning
  - inefficiencies
  - emergency hire
  - damaged items
  - moving in, out, about
- Additional capital
  - replacement equipment
  - new insurance spares
  - buildings and storage
- Consequential
  - penalty payments
  - lost future sales
  - legal fees
  - loss of future contracts
  - environmental clean-up
  - death and injury
- Administration
  - documents
  - purchase orders
  - meetings
  - planning and schedule changes
  - investigations and audits
  - invoicing and matching
  - utilities

ASSURING QUALITY IN MAINTENANCE

Alexander (Sandy) Dunn

Director, Assetivity Pty Ltd, PO Box 219, Como Western Australia 6952, Australia.

Abstract: For the purposes of this paper, we will define Maintenance "quality" as the extent to which maintenance actions lead to intended (desirable) outcomes. It is a central proposition of this paper, that maintenance quality is not currently being effectively or proactively managed. The latest research indicates that maintenance actions may be leading to desirable outcomes as little as 50% of the time. Yet most organisations have done little to actively measure, manage, or improve the quality of their maintenance activities. This paper will outline some practical actions that organisations can take to improve the quality of their maintenance activities. These include suggestions in the areas of:

- **Dealing with Human Error in Maintenance,**
- **Maintenance Task Design, and**
- **Maintenance Work Instructions and Documentation**

Key Words: Maintenance, Quality, Human Error, Task Design, Work Instructions, Quality Assurance

1 INTRODUCTION

Maintenance Quality Assurance can be considered in two ways. The first is at a higher, management level, which focuses on establishing the management processes, systems and performance measures that ensure that the maintenance function best meets the needs of the organisation. The second deals with Maintenance Quality at a task level – what actions must be taken to ensure that specific maintenance, repair and overhaul activities are performed to the required quality standard? This paper focuses on this second aspect of Maintenance Quality.

The assurance of Maintenance Quality is not a topic that, for the most part, has attracted much attention. Traditional approaches to Quality Assurance have revolved around either document control – particularly of maintenance procedures, and the timing and nature of predictive and preventive maintenance programs. These approaches have typically focused on ensuring that rules and procedures are followed – but, for the most part, with some notable exceptions, little effort has been put into ensuring that the maintenance procedures themselves support the delivery of high quality, consistent outcomes. In particular, maintenance task design, and the documentation of those tasks, rarely take into account the likelihood of human error, and even more rarely, design and document these tasks in such a way that human error is minimised.

2 WHY SHOULD WE BE CONCERNED ABOUT MAINTENANCE QUALITY ASSURANCE

There is an increasing body of evidence that suggests that poor quality maintenance should be a significant concern of all Maintenance Managers.

As far back as 1968, Nowlan and Heap [1], in their ground-breaking work that led to the establishment of the technique that we now know as Reliability Centred Maintenance, identified, when analysing the failures of hundreds of mechanical, structural and electrical aircraft components, that these failures occurred with 6 distinct patterns, as illustrated below.

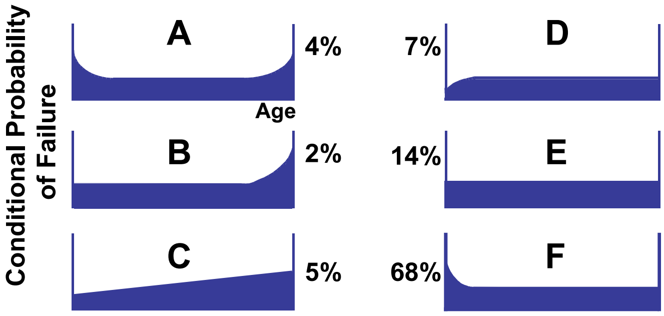


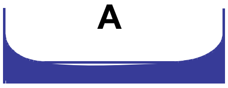
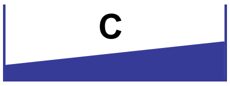
Figure 1 : Nowlan and Heap's Six Failure Patterns

The interesting finding, in the context of this paper, is that more than two-thirds of all components demonstrated early-life failure. It has been estimated that maintenance errors ranked second to only controlled flight into terrain accidents in causing onboard aircraft fatalities between 1982 and 1991 (despite the application of RCM techniques in the airline industry during this period) [2].

Subsequent studies have largely validated Nowlan and Heap's statistics. This includes work performed by Broberg in 1973, MSP in 1982 and SUBMEPP in 2001 as shown in Table 1 below

Table 1

Reliability Characteristic Studies

	Nowlan & Heap (1968)	Broberg (1973)	MSP (1982)	SUBMEPP (2001)
A 	4%	3%	3%	2%
B 	2%	1%	17%	10%
C 	5%	4%	3%	17%
D 	7%	11%	6%	9%
E 	14%	15%	42%	56%
F 	68%	66%	29%	6%

Reprinted from: Allen Timothy M., U.S. Navy Analysis of Submarine Maintenance Data and the Development of Age and Reliability Profiles, Department of the Navy - SUBMEPP, 2001, p.10

The Nowlan and Heap and Broberg studies were both conducted on aircraft, while the MSP and SUBMEPP studies were conducted on US Navy vessels. It should be noted in this table above that the low level of Failure Pattern F failures in the MSP and SUBMEPP data is explained by the fact that navy vessels go through extensive testing before they are returned to service following significant routine maintenance – any failures that are experienced during this testing period were not captured in the failure database used by the MSP and SUBMEPP projects [3].

A study of coal-fired power stations indicated that 56% of forced outages occur less than a week after a planned or maintenance shutdown [4].

Professor James Reason [5] has compiled a table summarising the results of three surveys – two performed by the Institute of Nuclear Power Operations (INPO) in the USA, and one by the Central Research Institute for the Electrical Power Industry (CRIEPI) in Japan. In all three of these studies, more than half of all identified performance problems were associated with maintenance, calibration and testing activities. In comparison, on average only 16% of problems occurred while these power stations were operating under normal conditions.

Reason also quoted the results of a Boeing Study [6] which indicated that the top seven causes of in-flight engine shutdowns (IFSDs) in Boeing aircraft were as follows:

- Incomplete installation (33%)
- Damaged on installation (14.5%)
- Improper installation (11%)
- Equipment not installed or missing (11%)
- Foreign Object Damage (6.5%)
- Improper fault isolation, inspection, test (6%)
- Equipment not activated or deactivated (4%)

We can see from this, that only one of these causes was unrelated to maintenance activities, and that maintenance activities contributed to at least 80% of all IFSDs.

A study conducted by Boeing and US ATA [7] found that maintenance error was a crucial factor in aircraft accidents from 1982 to 1991, contributing to 15 % of the commercial hull loss accidents where five or more people were killed. Rankin and Allen [8] established the economic costs of these maintenance errors, estimating that 20 to 30 % of in-flight shutdowns are due to maintenance error, 50 % of flight delays are due to engine problems caused by maintenance errors, and 50 % of flight cancellations are due to engine problems caused by maintenance errors.

If poor quality maintenance causes so many incidents in highly regulated and hazardous industries such as Nuclear Power Generation and Civil Aviation, what proportion of failures may be being caused by Maintenance within organisations that operate in less regulated environments?

What are the outcomes of maintenance-induced failures? Clearly, depending on the industry in which you operate, there are potentially significant safety and environmental risks. There is a long list of catastrophic failures in which, the inadequate performance of a maintenance task played a significant role. Some of these include:

- Flixborough
- Three Mile Island
- Piper Alpha
- American Airlines Flight 191
- Bhopal
- Japan Airlines Flight 123
- Clapham Junction
- Etc. etc.

But besides the obvious safety risks, perhaps the bigger consequences are economic. General Electric has estimated that each in-flight engine shutdown costs airlines in the region of US$500,000. Replicate that figure across other industries, and the cost of poor maintenance quality could be staggering?

Clearly, we need to do something to reduce the number of equipment failures that are being caused, not prevented, by maintenance. Yet this area has received very little serious attention by most organisations.

3 DEALING WITH HUMAN ERROR IN MAINTENANCE

Reason [9] argues strongly that human error is inevitable. It is an unavoidable consequence of our psychological and physiological make-up. He argues further, and convincingly, that error is not limited to a few "high risk" individuals, but is universally spread throughout the population – that sometimes the worst mistakes can be made by senior people carrying the greatest responsibility. If we accept this proposition, then since it is difficult to eliminate errors completely, continuing emphasis must be placed on developing interventions to either reduce the likelihood of human error in maintenance, or to make maintenance procedures more error-tolerant.

Reason and Hobbs [10] identified several risk areas where maintenance error is most likely to occur. We can summarise these risk areas by categorising them according to whether they relate to individuals, tasks, workplace supervision and organisation, or the enterprise.

- INDIVIDUALS
  - Reliance on Memory – the greater that an individual tends to rely on memory – particularly for more complex tasks – the greater the risk that the task will not be performed correctly
  - Pressures – for more complex tasks, the greater the time pressures placed on the individual (either direct or implied) the greater the risk of an error being made
  - Tiredness – the more fatigued the individual, the greater the risk of an error being made. This fatigue is often a direct result of shift rosters and patterns, as well as circadian cycles when working shift work.
  - Competence at Performing Tasks – the less competent, or familiar, an individual is with a particular task, then the greater the risk of error
  - Deliberate Violations Based on Individual's Risk Assessment – some individuals are more likely to demonstrate high-risk behaviours by not conforming with specified rules and procedures

- TASKS
    - Ambiguity of Task – if the task to be performed is not clearly specified, and the procedure to be followed in performing the task is not clearly understood, then there is a higher risk of error
    - Frequent, Repetitive, Routine Tasks – tasks that are performed frequently by the same individual, and which require low levels of attention in order to perform those tasks represent a higher risk of error
    - Design Issues Affecting Maintainability – if equipment is not designed ergonomically to permit adequate access and lighting to perform maintenance tasks, or increases the level of fatigue when performing maintenance tasks, then there is greater risk of an error being made.
    - Fatigue – if the task to be performed is strenuous, or requires considerable physical or mental effort, then there is a higher risk of individual fatigue, and resulting error.
    - Past Errors – if a particular task or task step has been performed incorrectly in the past, this is an indication that this particular task is likely to be more prone to future error.
    - Installation Tasks – the risks associated with error are greater for installation or assembly tasks than they are for removal or disassembly tasks.
    - Complexity of installation/assembly – if the task is more complex
    - Functionally Isolated Tasks – steps in a procedure which are not obviously cued by preceding actions are more likely to be subject to error.
    - Recent Changes – tasks or procedures which have recently been changed – particularly those which have been performed many times previously using an old procedure or method, are more likely to be subject to error.
    - Differences with Similar Tasks – tasks which are very similar to other tasks, but have few, but important differences, are more likely to be subject to error
    - Hidden Errors – if a task or step in a procedure was omitted, or performed incorrectly, and this error is subsequently hidden from view, then this represents a greater risk
    - Visibility/availability of cues – if the task required cues, parts or tools that are not readily at hand and visible, then there is a greater risk of error
    - Close to Completion – if the task is performed close to the end of a larger activity, then there is a greater level of risk. This is particularly the case if the task is to occur after completion of the main objective, but before the entire activity is complete.
    - Multiple Personnel Involved in Completing the Task – if the task is to be performed by more than one person – and particularly if task completion requires handover to another person – then the risk of error is greatest.
    - Require Removal of Tools or Other Items – if, on completion of the task, tools, parts, or other items are required to be removed from the equipment or the immediate location of the task in order to permit safe and reliable operation of the equipment, then the risk of error is greater
    - Optional Steps – if the task includes optional steps that are sometimes required, and at other times not required, then the risk of error is greater.
- WORKPLACE, SUPERVISION AND ORGANISATION
    - Availability of Supervision – if supervisors and team leaders are not available to perform regular, on-the-job assessments of work quality, and compliance with procedures, there is a greater risk of an error
    - Quality of Monitoring and Supervision – if supervisors do not adequately accept responsibility for work quality, do not perform regular, on-the-job assessments of work quality, and do not monitor compliance with procedures, there is a greater risk of an error
    - Reliance on Remembering Complex Instructions – if, when individuals are allocated work, they are expected to remember complex instructions, then the risk of error is greater.
    - Interruptions – the more interruptions that an individual is subjected to when performing a task, the more likely that an error will occur – usually an error of omission.
    - Breakdowns in Coordination – tasks that require coordination between individuals and groups – particularly those tasks which are started by one individual, and continued by another, as is frequently the case with tasks that take more than one shift to complete – represent a higher risk of error
    - Excessive Authoritarian Leadership Style – leadership styles which do not recognise the inevitability of human error, and which respond to such errors with punishment, rather than seeking understanding, and

searching for practical solutions to avoid recurrence are more likely to lead to errors being hidden – the net results being failure to reduce the number of errors occurring

  - Non-compliances based on Group Norms - some work teams and groups are more likely to demonstrate high-risk behaviours by not conforming with specified procedures, as a result of peer group norms that have been established
  - Fatigue – the more fatigued the individual, the greater the risk of an error being made. This fatigue can be related to the nature of the work that has been assigned to individuals, and the length of the shift worked – particularly if overtime is involved.
  - Poor Housekeeping – untidy and cluttered workspaces in the area where tasks are being performed are more likely to lead to parts or tools being misplaced, or the incorrect part or tool being used.
  - Inadequate Spares – if spare parts are not readily available for important maintenance tasks, then there is a risk that alternative, less reliable parts will be used, or that parts that should be repaired or replaced will be reused. This represents an increased risk of failure.
  - Inadequate Tools and Equipment – if the correct tools and equipment required in order to perform a task properly are not available, then there is an increased risk that alternative tools or methods will be used, with consequent increase in the risk of an error.
  - Shift handovers – inadequate communication at shift handover, where a task is to be continued on a following shift, represent significant risk. If the status of the current task is not adequately communicated, then there is a risk of a critical step or action being omitted.

- ENTERPRISE-WIDE
  - Non-compliances based on Perceived Cultural Norms – some organisations and departments are more likely to deviate from approved procedures, as a result of explicit or implicit sanctioning of such behaviours.
  - Failures to Detect or Respond to Non-Compliances – if there are inadequate formal, independent quality assurance processes in place, there is greater risk of an error
  - Shift Rosters – the more fatigued the individual, the greater the risk of an error being made. This fatigue is often a direct result of shift rosters and patterns, as well as circadian cycles when working shift work.
  - Organisational Structure Fails to Support Elimination of Errors – if accountability for monitoring, and improving, work quality is not clear, or is not adequately resourced, then there is increased risk of errors
  - Ineffective People Management – if individuals are not explicitly made aware of their role in ensuring adequate work quality, then the risk of error is increased
  - Ineffective Training and Selection Process – if competencies are not adequately assessed, and improved, the risk of errors is increased
  - Commercial Pressures – organisations where commercial pressures for short term financial performance are extreme are more likely to be subject to errors
  - Operational Pressures – organisations where pressures to achieve short term operational output targets are extreme are more likely to be subject to errors
  - Inadequate Planning and Scheduling Processes – organisations where operational and maintenance planning and scheduling processes are not sufficiently well established, or do not plan or schedule work with adequate detail, are more likely to be subject to errors
  - Inadequate Maintenance and Warehouse Infrastructure – organisations where maintenance workshop facilities and warehouse facilities are inadequate, or inadequately maintained, are more likely to be subject to errors.
  - Inadequate Error-Detection and Error-Containment Programs – organisations which do not have independent quality assurance processes in place to detect, and correct or contain the effects of errors are at higher risk.

So how can we deal with this multitude of risks that can contribute to maintenance error, and a loss of task quality? Reason and Hobbs suggest several different types of interventions.

For dealing with the individual risks, Reason and Hobbs suggest that two key activities take place. The first of these is training individuals in error awareness. This is intended to make participants aware of the factors that are most likely to lead to them making errors, to recognise when these factors exist, and provide them with strategies and methods of dealing with these factors. The second activity is aimed at reducing the number of deliberate violations performed by individuals. This means dealing with the three factors that impact on the likelihood of individuals to deliberately break the rules – their individual attitudes to risk-taking behaviour, social controls that individuals take into account when deciding whether to break the rules, and the extent to which individuals feel that they can control and modify their own behaviour.

For task-related risks, there are several alternative proactive measures that can be taken. These include:

- Effective equipment design for maintainability
- Designing the task to minimise the situations in which errors are most likely
- Using effective Maintenance Work Instructions and documentation

We will discuss the last two of these in more detail in following sections of this paper.

For dealing with workplace, supervision and organisation risks, there are also several actions that can be taken, including:

- Minimising individual fatigue through effective work allocation
- Effective allocation of work to individuals to minimise repetition and boredom, as well as ensuring that competent people are allocated to work
- Structuring workplaces to minimise the likelihood of interruptions
- Developing strong processes for coordination within and between work teams – particularly shift handover processes
- Ensuring that work is well planned and scheduled, that adequate time has been allocated for completion of the activity, and that all required parts, tools and other equipment are readily available
- Ensuring a tidy workplace, and encouraging tradespeople to maintain organised and tidy work environments while they are performing tasks, as well as after they have completed the task.

At an organisational level, Reason and Hobbs recommend the implementation of both reactive and proactive measures for dealing with errors.

Reactive measures could include formal Root Cause Analysis of appropriate incidents after they have occurred. This Root Cause Analysis process, to be effective should fully investigate all of the contributing causes to the failure, whether these be physical causes, human causes, or organisational causes. The most effective solutions to preventing these failures from happening again, will be those that deal effectively with the organisational causes of failures.

Proactive measures could include formal risk assessments, investigations of "near misses", and the proactive identifications of gaps in the organisational defences that reduce the risks associated with maintenance errors. Possible areas that could be assessed in this risk assessment would include:

- The knowledge, skills and experience of maintenance personnel at all levels
- Employee morale
- The availability of tools, equipment and parts to perform maintenance tasks
- Workforce fatigue, stress and time pressures
- Shift rosters
- The adequacy of maintenance procedures and work instructions

Reason and Hobbs consider that effective organisational measures to minimise the risk of maintenance errors require the establishment of a "Just Culture", "Reporting Culture" and "Learning Culture" within the organisation. A "Just Culture" is one where there is a strong level of trust between all individuals and groups within the organisation. They see this as being the prerequisite to the establishment of a strong "Reporting Culture". A "Reporting Culture" is one where all errors and failures, no matter how seemingly insignificant, are reported. A "Learning Culture" is one where the lessons of the past are remembered, and solutions are effectively implemented in order to minimise the risk of future errors. Those who have researched so-called "High Reliability Organisations" (HROs) have noted that high levels of failure reporting are a significant feature of those organisations [11].

4 MAINTENANCE TASK DESIGN

In the area of task design, there are many things that can be done to ensure that the risks of unintended consequences are minimised. These could include:

- Ensuring that adequate "hold" points are put in place at critical points in maintenance tasks, beyond which point tradespeople should not proceed without having their work checked by an independent person, and the quality of the work performed to that point verified. This could particularly be the case where the work performed may not be visible if the tradesperson was to continue. Examples of this could include inspection of internal vessel, turbine or heat exchanger repairs after the work has been completed, but before the equipment is closed up and returned to service.

- Incorporating adequate reminders within job procedures which are more likely to make omissions more obvious. For example, if there are items which must be removed in order to conduct maintenance repairs, and which must be replaced before the equipment is returned to service, and replacement of these items has a high risk of being overlooked, then these items may be tagged with some form of highly visible flag in order to make their non-replacement more obvious.
- Ensuring that parts which are removed and which are to be replaced and not refitted are placed immediately in a different location to the new parts which are to replace them. Parts may even be marked or tagged to indicate that they are to be scrapped or refurbished.
- Encouraging tradespeople to place parts that have been removed, and which are to be refitted, in an ordered or structured arrangement – so that all parts to be refitted can easily be identified, and failure to refit a required part is more easily identified. For regularly performed jobs, a kit or shadow board may be used to temporarily store these parts while repairs are being performed.
- Another alternative, to ensure that parts that have been temporarily removed are refitted is to secure these items to the equipment by means of a tether. This prevents these parts from being moved to a location where they may be overlooked.
- Ensuring that items of machinery are adequately protected from contamination by the ingress of dust or other foreign objects into sensitive internal parts while these machines are being maintained. This could take the form of fitted covers over inlet points.
- Ensuring that work is performed in logical sequences of grouped actions. The risk of error is higher if the activities performed are not logically related.
- Attempting to minimise the number of tasks to be performed after completion of the main objective of the maintenance activity. Where possible, these tasks should be sequenced to occur prior to the main objective being achieved.
- Include, where practical, full functional testing of equipment prior to it being returned to service.

5 MAINTENANCE WORK INSTRUCTIONS AND DOCUMENTATION

The quality and content of Maintenance Work Instructions and Documentation is vital, as evidenced by the following findings from the civil aviation industry [12]:

- A 2004 UK Flight Safety Committee determined that the top three causes of maintenance mishaps were:
  - Failure to follow published technical documentation
  - Using an unauthorized procedure not referenced in the technical documentation
  - Supervisory personnel accepting non-use of technical documentation or failure to follow maintenance instructions.
- A 2001 FAA study found that documentation is the leading contributing factor to maintenance events. In most cases organizational issues resulted in failure to use documentation that was available.
- A 2002 study of procedural issues from NASA Aviation Safety Reporting System incident reports showed that the following factors contributed to documentation-related errors: procedural design flaws, user errors, currency and accessibility of documents, and organizational practices.
- Aviation maintenance personnel spend from 25-40% of their time searching for, using and completing maintenance documentation.

The Federal Aviation Administration includes the following advice regarding actions that should be taken to ensure that Maintenance Work Instructions and Documentation are of adequate quality to minimise the risk of poor documentation leading to maintenance error [12]:

- Develop guidelines and standards for documentation development. The guidelines/standards should include the following:
  - Use a standard format and page layout.
  - Use one action per statement.

- Write DANGER, WARNING and CAUTION statements using current international (American National Standards Institute/International Standards Organization) standards.
    - Avoid using acronyms and abbreviations.
    - Ensure that where an arrow is pointing to a figure or graphic it is immediately obvious.
    - Use (…) Simplified English.
    - Verify the documentation by having a second person check for technical accuracy.
    - Validate the documentation by having a sample of users check for usability.
- Provide a process for immediate and longer-term improvements to technical documents based on user feedback.
- Ensure prompt acknowledgement to submissions for document change.
- Respond to the document improvement suggestion in a timely manner.
- Develop measures to track specific documentation problems and corrective actions.
- Be sure the document delivery system is functional and well maintained.
- Use advice from documentation research and industry best practices.

Additional steps that should be taken to ensure that maintenance work instructions and documentation minimise the risk of error (based on information contained in [13] and [14]) include:

- Ensuring that illustrations are used where appropriate, as these can often convey instructions far more quickly than written text.
- Ensuring that any illustrations used are clear, and correct
- Ensuring that all steps required in order to perform the task have been included within the procedure
- Ensuring that work instructions and documentation are readily available to the person performing the task.
- Ensuring that all instructions required in order to perform the task are contained within the one document –references to other essential documents should be avoided wherever possible
- Ensuring that the work instruction is accurate for the specific make/model/serial number of the equipment being worked on. Frequently "generic" work instructions do not apply to specific equipment items – particularly if these have been modified.
- Ensuring that all the steps are in the correct sequence
- Ensuring that similar procedures in different references agree (for example, that the work instruction given on a job card agrees with that contained in the Manufacturer's Manual, or vice versa).
- Ensuring that the procedure can be followed, in practice
- Ensuring that there is proper version control over work instructions and job procedures
- Group complex work instructions into phases, with each phase consisting of many, related tasks
- Focus on the key risks that may prevent the job from being performed safely and to the required quality standard.
- Use call-out boxes and other methods to highlight key actions that must be taken, tolerances that must be met and other items that represent the highest risk of the task not being performed correctly
- Ensure that instructions are written in the first person, not the third, and use the active voice, not the passive.

6 CONCLUSION

This paper has attempted to illustrate that the issue of Maintenance Quality is one that is not currently receiving adequate attention within most organisations. It has also attempted to illustrate that improving Maintenance Quality has the potential to generate significant improvements in organisations' financial, safety and environmental performance. It went on to discuss some specific risk areas relating to maintenance error, and make suggestions of actions that could be taken in designing maintenance tasks, and documenting work instructions, in order to minimise the risk of maintenance error, and enhance Maintenance Quality.

7 REFERENCES

[1] Nowlan FS & Heap H (1978) *Reliability-centered Maintenance*, Springfield, Virginia: National Technical Information Service, US Department of Commerce.

[2] Davis RA (1993) – *Human Factors in the Global Marketplace* – Keynote address, Annual Meeting of the Human Factors and Ergonomics Society, Seattle, 12 October 1993

[3] Allen Timothy M., *U.S. Navy Analysis of Submarine Maintenance Data and the Development of Age and Reliability Profiles*, Department of the Navy - SUBMEPP, 2001

[4] Smith A (1992) – *Reliability Centered Maintenance* – Boston, McGraw Hill,

[5] Reason J (1997) – *Managing the Risks of Organizational Accidents* – Aldershot, Ashgate Publishing,

[6] Boeing (1994) – *Maintenance Error Decision Aid*, Seattle: Boeing Commercial Airplane Group

[7] Reason J (1997) – *ibid*,

[8] Boeing/ ATA, (1995) – *Industry Maintenance Event Review Team*, The Boeing Company, Seattle, WA.

[9] Rankin, W.L., and Allen, J., (1995) – *Use of the Maintenance Error Decision Aid (MEDA) to Enhance Safety and Reliability and Reduce Costs in the Commercial Aviation Industry*, Proceedings of the International Air Transport Association's 1995 Aircraft Maintenance Seminar and Exhibition "The Changing Vision", November 14-16, Sydney Convention and Exhibition Centre, Sydney, Australia.

[10] Reason J & Hobbs A (2003) – *Managing Maintenance Error*, Aldershot, Ashgate Publishing, 2003

[11] Weick, Karl E. & Sutcliffe, Kathleen M. (2001) – *Managing the Unexpected – Assuring High Performance in an Age of Complexity* – San Francisco Jossey-Bass

[12] Federal Aviation Administration (2005) – *Operator's Manual - Human Factors in Aviation Maintenance* - found at http://www.hf.faa.gov/opsmanual/assets/pdfs/HFOM_Maint_Org.pdf

[13] Boeing (1994) – Maintenance *Error Decision Aid (MEDA) Users Guide*©, Seattle: Boeing Commercial Airplane Group

[14] Dunn A (2004) - *Managing Human Error in Maintenance* – Proceedings of MESA WA's 2004 Maintwest conference, Technology Park, Bentley, Western Australia

ALIGNING OPERATIONS, MAINTENANCE AND ENGINEERING TO PROVIDE THE SEEDS OF SUSTAINABLE ASSET MANAGEMENT – A CASE STUDY

G Williamson

Assetivity Pty Ltd., Suite 12, Como Corporate Centre, 11 Preston Street, Como, Western Australia, 6152, Australia..

Abstract: Successfully delivering large capital projects imposes many challenges on both Project Management and the Operation and Maintenance teams. This paper discusses some of the key issues relating to the Projects/Operation's and Maintenance interface, as experienced by the author while leading a large, asset-intensive organisation during a period of enormous growth. The paper will share the learning experiences gained not only through the development of this organisation but also some of the technical issues which impact on plant and asset management performance during the early life of a major new asset. The paper provides an insight, from an Asset based Operation's Department perspective, of how to improve Engineering, Operation's, Maintenance and Project Management processes to realise not only cost and schedule advantages but also to improve post start up performance of the new equipment.

Key Words: Project management, operatons, maintenance

1 INTRODUCTION

Sustainable Asset Management is not just related to the physical asset and how we engineer, operate and maintain those assets. To be successful we must take a holistic approach by taking into account all the aspects which impact asset management including the 5 broad areas of (1) Operation's and Maintenance Strategy, (2) People, Skills and Culture, (3) Organisation Structure, (4) Management Processes and Practices, (5) Use of Technology. This paper will explore how all these factors were attempted to be addressed and the learning achieved for both the Operating and Project organisations. The case study is derived from the authors own experiences and reflects the Capital and Operating intensive environment of the period 1997 through to the end of 2005, within the LNG area of the Oil and Gas industry.

Throughout the paper the term Operation's or Operation's department's refers to the structure of the company in which the author worked, and is the collective term for the Production, Maintenance and Engineering Departments which formed the Operation's Group.

The Management of Assets are also not immune from the particular business environment any Asset Holder may find themselves in. What may be considered Good, Excellent or Acceptable Asset Management in one business area, may be totally different to the business requirements of another area. Also Life Cycle planning for given assets may differ across industries, with expected life spans being anything from a small number of years to much longer periods.

With many variables placed before those charged with the task of Asset Management, it is clear that we must start by defining as many of them as possible. Therefore this paper and its supporting presentation starts by explaining the Business environment of the case study, this is done to provide a basis for the reader to understand the subsequent sections, and also to assist in comparing their own circumstances, and perhaps provide some ideas, in pursuit of Sustainable Asset Management in their own unique business areas. In Life Cycle Terms the paper focuses on the development, start up, and initial operation phases and provides the reader with an insight of the Learning Experiences gained by Operations and Projects within this part of the Asset Life Cycle

2 CAPITAL PROJECT AND OPERATION'S EXCELLENCE ~ A CHALLENGE TO ALL

Within a short period of time, namely three years, the company had experienced the establishment of its production assets and people assets in a Greenfield environment and had witnessed the transition into a full production company. A further two years were to pass during which time the benefits of a company and its resources growing together, were already providing rewards in terms of better than expected availability figures. Nevertheless the competitive environment the company found itself in was requiring management to improve business performance. This resulted in the need to compare performance through industry benchmarking and to provide strategic targets to the organisation to enable new sales and development to take place. LNG markets through out the world during the period 2000 – 2005 were characterised by demand outstripping supply and therefore new production facilities were required. This challenged the technical design teams to provide even larger plant capacities at a stable or reducing Capital investment cost. To fund these projects and also to demonstrate sound production management, the Operating and Maintaining organisations were required to meet high levels of availability, to demonstrate reliability of supply to customers at the same time doing so at a reduced cost of operations. The author found himself in this environment initially as Maintenance Manager then transitioning through the role of Manufacturing Manager to a totally new function as Operations Expansion Manager. These positions developing as a result of the pursuit of lower or best in class operating costs, in an environment of enormous expansion and capital investment to satisfy the LNG industry and its customer's demands for secure and reliable supplies of environmentally friendly energy sources.

The challenge facing the author and the company only two years after initial operation was to achieve class leading production capability at lowest cost, while working in an expansion environment of some of the largest ever capital developments, in the Oil and Gas sector.

The establishment of the first two trains in the period 1996 ~ 2001 was fully supported by the Operation's department resources, during this time the site work progressed as the company established its Operating Organization and recruited and developed its staff and their competencies. In terms of Operations involvement the entire department was able to focus and support the Project as well as growing in learning as the work progressed towards commissioning and start up. It should be noted that following successful start ups, the first two Trains have achieved world class reliability.

During the 2001 to 2003 period as the Operation settled into the routine production plan, it became opportune to look externally at similar Operations and see what lessons could be learned and put in place to improve work processes with a target of minimizing Operating costs. At the same time however, new larger production facilities were being planned and contracts established for Engineering and Construction of these facilities, with the Project organisations being tasked with finding improvements in Capital costs and project schedules. A clear concern quickly developed in this scenario regarding the ability to provide the people and processes from the Operation's departments to support involvement in the new projects, and of course to ensure that sustainable Asset Management principles were upheld.

As a result of potentially conflicting business directions it was recognized that learning, and team work, would play a vital role in the successful start up of new facilities. By early 2004 with the announcements of even more increases in production capacities, Operation's involvement in projects became critical to ensure the company could meet the expectations of its customers. From an original capacity of 1.2 Billion SCFD the company would see growth to 3.6 Billion SCFD by 2005, with a current capital investment portfolio taking the company to over 10 Billion SCFD over the following five years.

We will now take the reader through some of the innovations the Operation's departments have made to the way they work and learn from experience in a dynamic capital intensive environment.

3 A STORY OF LEARNING IN THE OPERATION'S DEPARTMENTS

In reviewing the direction taken along the path of improvement in the initials years, it is important to refer to some first principles. The first stage therefore on this journey is to address the term "Learning".

In a busy environment of Engineering, Constructing and Commissioning a project, the demands faced by the project departments and the Operations departments can sometimes be seen as non related. This is of course not the case as the operability and maintainability input which the Operations Department can offer to the various stages of the project life cycle can indeed have benefits to both organisations. The first lesson we would offer, however, is that Learning will not just happen and be of future benefit, unless there is a structure put in place by which lessons can be captured, recorded, analyzed, and eventually used to improve current and future projects.

Several techniques can be used to ensure that good asset management lessons can be captured, and the traditional technique of post project workshops was the tool selected at the end of the 1990's by the Operations and Project departments to capture those lessons. These were held post start up and resulted in the creation of an extensive, and most importantly, agreed list of lessons to be applied in future projects. This practice continued through the period when the third train became a firm project. There were, as would be expected in any company wishing to promote continuous improvement, some enhancements to the learning process. First it must be noted that the list of lessons derived from the experience of the initial two trains, was reviewed, and where appropriate, applied to the new project. The structure of how the company approached the learning processes however changed significantly. Obviously lessons still had to be captured, therefore use was still made of the workshop structure, but now the workshops were tailored to specific areas, such as Construction Activities, Start Up activities, engineering activities etc. The most important change however was the establishment of an on line Continuous Improvement Data base.

The data base was structured as a repository for several key content areas. The first of these areas was Technical Queries where items can be captured at any phase of the project and by anyone involved in the project. The database provided an audit trail to ensure that resolutions were worked to eliminate any problems. The second area was warranty items which also appear pre and post start up, and may lead to improvements being made to future projects. This would occur once the claim is analysed and rectified with the potential root cause of the problem which resulted in the claim, being identified and understood. Of course not all items will lead to a new lesson for Operation's, as they may result from manufacturing defects or installation defects, but even in these areas learning observations can still prove beneficial to future work. The third and obvious key area was the creation of a central register of all the lessons learned which could be accessed by all, and will grow as more lessons are developed, becoming an invaluable source of data.

In establishing and using the data base a realized hidden benefit appeared in terms of the ability of the data to be interrogated, similar to a Root Cause Analysis tool. As an example, by looking at the various areas such as Technical issues and Warrantee issues we were able to identify a trend of QA/QC issues in the manufacturing processes of various suppliers, which required enhancement of quality control support to hopefully benefit future projects. A key feature of using a data base which is accessible to all relevant parties and can be searched by multiple criteria, was it not only captured Learning, but we were able to put it to far greater use.

To make the Continuous Data Base a sustainable tool it was required that each function nominate a gatekeeper of data for their own areas of business. Therefore specific functions within both the Project Organisation and the Operations organisation nominated such gatekeepers. The role of the gatekeeper was first to be the custodian of improvement data input to the database and along with other gatekeepers, being responsible for database improvement. With the second key function of being an advisor to their respective managers, as to how the database was being utilised by their function or department

4 INFLUENCES WHICH SHAPED OPERATION'S WORK

With large complex projects, ensuring that everyone involved understands the roles and responsibilities each have, can lead to some surprises and gaps even in the most well thought out projects.

During the Commissioning and Handover of the third train project to Operations for start up, it was found that there were several areas where each party thought the other was clear on who did what and when. This was the case with the completion of Hydrostatic testing of pipe work, where the EPC contract clearly indicated that Operations were responsible for re-instatement of pipe work. However, because of the lack of attention to detail and also the assumption that specifications would be the same as the previous plant, the issue was missed. Although at the time the resolution was worked and mutual support provided for the benefit of the project, it was clear that a gap had appeared. One of the learning points here was the need to treat every situation as unique and to ideally ensure that changes to practices and specifications within a contract document also go through the same Management of Change process that changes to the actual facilities do.

This resulted in the need for some analysis and focus on the complete picture related to Handover and Commissioning, and resulted in the introduction of clearer process flow charts of this stage of the project, and amended contracts where appropriate. This provided clarity in contractual boundaries as well as roles and responsibilities concerning Witnessing and Verification activities.

With the significant interaction between several projects at differing development stages and the use of synergies between them and the running revenue generating plant, to limit costs of the development. There emerged a significant requirement for Tie Ins of new equipment to existing running plant. These tie ins were not just restricted to mechanical equipment such as pipe work, but also involved Electrical and Control Systems tie ins including software programme changes. With the list of projects facing the company, and also the time span between the purchasing of the original plant and the new project, it was clear that this area was going to command a lot of attention. This manifested itself for example, where the latest version or model of equipment was to be used in a new project, but relied on some functionality of plant purchased up to 10 years previously, and then inevitably the older plant systems would have to be upgraded. The upgrading however had to be accommodated within the commercial business boundaries of the running equipment. To approach this work it was decided to establish a full time role for the coordination of all the activities associated with Tie Ins and the impact on the execution of the company's Management of Change processes. Resources were dedicated to this task and the work load continuously monitored to ensure that correct procedures were applied in this critical area. This was aimed at ensuring that construction schedules were met, and also most importantly, that there was no impact on the running of existing equipment, through the potential for tripping running plant.

Allied with the technical queries mentioned in the previous paragraph, during the final stages of train three works a Doubts and Concerns Forum was established. The purpose of this forum was to answer questions raised by Operations staff as they became involved in the start up phase. Simple examples involved the placing of valves into the pipe work the correct way round, and more complex problems such as the running performance of pumps and compressors after initial introduction of product.

Another area was evidenced from Trains 3 and 4 work, namely the correlation between schedules. Given the size of these projects and the various geographical locations of design offices and manufacturers, it became clear that schedule dates could vary significantly in these different locations. To improve the schedule data a single scheduling tool was introduced which provided the capability for each individual schedule owner to manage their own data, but when rolled up into higher level company schedule, any conflicts would be immediately flagged and resolution sought. This of course could only be achieved by all parties using the same IT platform, which was a significant task in accommodating all locations including several major and minor contractors and suppliers of equipment.

As with all projects it was anticipated that equipment may fail during start-up and initial running due to design or manufacturing defects. The project staff in such situations is an invaluable support resource to Operations during this time.

From experience the signing of the handover mechanical acceptance paperwork did not necessarily mean the end of projects involvement. Therefore the need for an element of continuing project engineering support was identified and put in place as a permanent feature for future start ups.

One of the most critical changes Operations made post Train 3 experience was that an Operations organization which is geared to best in class operating structures would not be able to adequately contribute to the various project phases. To counter this, an entirely new function was formed, namely the Operations Expansion Department. The purpose of this new department was to provide the Operation's support by coordinating engineering resources to the early stages of the project, and also provision of a start up team. The Start Up team had the role of handling the handover processes and initial operation, eventually in turn passing care and custody to the running Asset Teams.

5 ASSET BASED ORGANISATIONS AND SUPPORT TO ASSET DEVELOPMENT

As previously mentioned the most significant change to the way Operations approached project work was the creation of a dedicated department to handle the project interfaces and the sound application of Operability and Maintainability requirements into future projects

One of the most aggressive and challenging learning's initiated in 2004, was to:

- Set up an Operations Expansion Team to address, support, coordination and resolution of all pre-commissioning, commission and start up issues
- Also the Operations Expansion Team was tasked to effectively interface with Operating and Engineering Experts to ensure that the best solutions to problems were identified and assessed, with the solutions satisfying the Asset Management principles of sound engineering practices, construction, and integrity specifications.

In the first year, the Operations Expansion Team, at a very minimum, relieved the existing Operations Asset Teams from the combination of handling daily problems as well as all the expansion/growth issues. This had an immediate positive impact allowing the Operation Asset Team to focus primarily on "running" the existing facilities. Whereas previously, the Expansion work load had been managed within the existing Asset Teams, which obviously proved to be an added extra stress to managing and coordinating an already full operations work plan.

As all operations teams find out, simplifying the daily job tasks and focusing on the key operation jobs allows the Operations Manager to reduce manpower to what is absolutely required for economic unit cost management.

Through the use of clearly mapped Project and Handover processes, the Operations Expansion Department were able to focus on meeting the challenges in its involvement with Projects. Further as the department carried the responsibility to handover eventually to the Asset Teams, a clear focus could be given to such areas as adequate documentation in particular in the area of spare parts management and input to the company IT systems.

6 TEAMWORK – THE REALITIES

Large scale projects will only succeed as a result of good team work between the various parties, and some of the links which have been put into place as a result of experience to date are discussed below.

The Operations Expansion Department plays a key role in ensuring a successful and most importantly a seamless transition through the various project phases. This is not just restricted to the start-up phase but also to the transfer of information during the design and detailed engineering phases. Relationships with the main Operations group were achieved through the adoption of clearly linked Business Objectives, which were continuously stewarded. Attendance at the regular departmental meetings ensured that all were kept up to date on developments, ensuring that mutual support between the two Operation's functions could be negotiated. This also ensured that the correct level of technical expertise could be made available in support of activities such as Drawing reviews, Model reviews Reliability and Maintainability Studies, and Hazops.

In terms of the relationship with the project teams, the Expansion Department had its own staff placed in the various Engineering offices around the world, able to respond to project questions during design stages or to seek further Operations Support as appropriate. These roles were seen as highly beneficial to all. Again as with the parent Operations organization the attendance at project meetings laid the foundation for good communication links. A clear learning which was put into practice was the use of an "Offsite" meeting, held between the project and the Operations Expansion Department on a Bi-Monthly basis. This meeting addressed strategic issues and identified opportunities for improvements which were prioritized and given to owners to pursue and put in place.

The projects covered both the upstream as well as the downstream facilities in an environment of multiple projects all at different stages of the project life cycle. In this business environment schedules play a vital role, a common platform for schedule management and clear ownership of the various parts was needed.

Within the multiple projects and suppliers environment, it is not unreasonable to expect that some may not perform, and therefore that failures may happen. Through use of the continuous improvement database it became possible to identify trends and potential weak areas, this is turn allowed the formation of contingency plans to assist in pre-empting potential trouble areas. Typical proactive measures were enhanced testing, provision of spares or extra support manpower. The lesson was to be prepared and examine “what if” scenarios, perhaps based upon analysis of each projects learning points and technical queries, to establish contingency plans.

A frequent characteristic during the exciting phase of start up is that when equipment does mal function Operation’s staff jump in to fix the problem. A clear lesson here was to create discipline within the organisation to follow a logical process chain of, when something goes wrong Operations Identify then Engineering along with Operations analyse and rectify. Valuable time and more importantly failure history can be lost in the enthusiasm and excitement of start up. Problems do appear, so ask the process questions first rather than the technical questions.

Re-iterating the Learning Story, learning opportunities will be lost if they are not documented. With the Continuous Improvement Data base it was hoped to institutionalize the capture of learning opportunities.

7 CLOSING REMARKS

The life cycle of Asset Management starts in the design phases of a project and grows through the equipment selection, construction and commissioning of a new facility. Frequently these phases are characterised by the lack of involvement of those who will be charged with the health of the asset for many years to come, and this may be as a result of complacency or even a deliberate policy. In this paper we have attempted, through the telling of a story of learning, to show how one company approached the issue. Therefore by focusing on processes, systems and organisational issues, the chances of a successful Life Cycle of Asset Management were improved.

IMPROVING THE FINANCIAL BOTTOM LINE IN HEALTHCARE WITH CMMS

Duncan Newsome [a] and Kevin Farrington [b]

[a] *Principle Consultant Possumworks, Varna St Waverley NSW 2024 Australia, and Director Corporate Services, Calvary Health Care Sydney Rocky Point Rd Kogarah NSW 2217, Australia.*

[b] *Director, 3pi International, Capalabar, QLD 4157, Australia.*

Abstract: Private Healthcare has always had an eye on the financial bottom line, generally through close management of staff hours per patient day. These days with decreasing budgets and increased service demand, CEO's in the Public Healthcare System are faced with similar requirements of meeting budget targets. In the Aged Care sector with tight margins and aggressive competition meeting budget is critical to survival. This paper addresses an overlooked area by most healthcare providers, which assists in reducing both the current year outgoings and the long-term return. While you would be hard pressed to consider the maintenance department as a revenue centre, when viewed from another perspective with appropriate use of an efficient CMMS properly structured and supported the maintenance department can be considered to be an additional source of revenue by:

- **Early appropriate intervention (preventative maintenance) reducing the risk of early retirement of expensive and often complex machinery**
- **Detailed tracking of expenses related to locations, equipment and cost centres**
- **The ability to triage (sort) appropriate requests from inappropriate requests**
- **Accurate reporting on cost per sq meter**
- **Improved productivity from internal and contract maintenance staff**
- **The ability to monitor contractor KPI's and enforce contract compliance**
- **In this paper, we will address how healthcare market sectors can benefit from the maintenance professional by reducing the bottom line and increasing revenue benefit.**

Key Words: CMMS, Productivity, Savings, Healthcare

1 CURRENT ENVIRONMENT – HEALTHCARE MAINTENANCE

Private Healthcare has always had an eye on the financial bottom line, generally through close management of staff hours per patient day. These days with decreasing budgets and increased service demand, CEO's in the Public Healthcare System are faced with similar requirements of meeting budget targets. In the Aged Care sector with tight margins and aggressive competition meeting budget is critical to industry survival.

The current reality for maintenance within the Healthcare environment is a reduction in on site trade and supervisory staff to a level where only the most rudimentary repairs and maintenance is undertaken by employed trades staff, and the majority of Preventative Maintenance (PM) and repairs are outsourced. This is exemplified by the reduction over the past twenty years in employed maintenance trade qualified staff in most facilities. An average maintenance department for a medium sized facility of 100 bed capacity would a one time have had a maintenance department consisting of a maintenance supervisor and 3 – 5 trade qualified staff (plumbers, electricians, painters and chippies). Today these facilities usually have one trade qualified staff members who is responsible for the coordination of the PM and repairs program for the facility under the direction of the Director of Finance or another executive of the facility.

The maintenance budget is usually the last to be increased (if at all) to cover the normal increased cost provision of services to increasing complex buildings and assets. When a facility anticipates financial pressures to reduce cost, maintenance tends to be seen as an area in which cash flow can be instantly reduced to minimise anticipated budget deficits. This short term approach has little or no regard for the long term consequences resulting from the delayed or cancelled facility and asset maintenance which may occur if the budget situation continues for a number of years. This is of course at odds with the expectation that maintenance departments will maximise asset life, and the facilities will always look presentable to the customer (patients, relatives, etc). Typically failing infrastructure is always more expensive to fix at a latter point in the life cycle and if the delay is long enough the cost of repair or replacement will have increased exponentially.

As Dave Berger (2003) noted in his article on Asset Management – controlling the life cycle it is a "pay me now, or pay me more later situation". Effective use of a capital asset plan coupled with a good CMMS will ensure that money is saved and not wasted. It is within this environment that staff responsible for facility maintenance must deliver on the expectations of the owners or the executive. The key to harnessing the complete benefit to the organisation is by taking full advantage of the functionality available within the CMMS. The first step is to discuss facility maintenance with the facility directors in financial terms. It is imperative that, maintenance management is able to clearly articulate the benefits to the organisation, which will be derived from maintenance supported by an effective Computerised Maintenance Management System (CMMS). By addressing the following questions with management, the maintenance manger will have the framework to produce an effective and efficient delivery of PM and repair maintenance to the facility:

1. What is the goal of our CMMS implementation?
2. How will our CMMS assist us? And
3. What are our real expectations following implementation of the CMMS?

The revenue saved by maintenance by following a proactive maintenance framework, will improve the financial bottom line for that facility. Although, it needs to be noted that in change management terms, when implementing any program or project, there is a cost involved in terms of gaining commitment from the staff and the organisation or there is a cost from resistance to implement. Either way there is cost and effort involved if your project is to succeed.

2 PREVENTITIVE MAINTENANCE

While PM is a concept generally well understood in management and maintenance groups, the appropriate application and implementation of PM is not as well understood. Following installation of many key assets, such as lifts, HVAC and hydraulics, the new owners are presented with either: a comprehensive and expensive list of PM instructions from the installer who is also contracted to provide the PM & repair servicing to that asset item or an extremely sketchy list with some vague references to manufactures documentation. Unless the owners (either in management or maintenance) have the appropriate technical discipline to assess the presented PM schedule it is usually accepted as is. Significant external factors such as seasonal load variation are not often considered, and periodic PM occurs regardless of the asset workload. The cost implication in either case can be a significant impact on a facilities budget through either funding "overkill" PM program or early capex asset replacement due to failure of the item through under-maintaining. A classic example of this is High Volume Air Conditioning (HVAC) PM. Some of the external factors impacting on the asset which should be considered include: climatic conditions, particle contamination levels of intake air, building occupation rates; all of which vary throughout the year, thus rendering the standard PM routine of 3 monthly, 6 monthly and annual inspections / services as inappropriate to load. The important factor here is that generally the manufacture's recommended PM routine is based on the asset working at maximum capacity in the most testing conditions all year round.

It is crucial to know if the PM schedule is delivering the appropriate service for the asset to reach it's maximum life and producing any mandatory certifications required for the asset. Without an appropriately configured CMMS, it is a difficult task to achieve. If a facility is currently using a spreadsheet or hardcopy, which can be entered into a spreadsheet, all the facility data could be transferred into the CMMS with the set-up defining frequencies and standards for each frequency. At a predefined interval the CMMS will print out a work order with the predefined tasks, job steps and any Safe Working Practices (SWP) associated with the service. The CMMS should adjust the next date for service dependant on the date the last service was completed. The savings from having PM's automated using a CMMS result from a lesser clerical load on the maintenance staff, and by not having earlier than necessary servicing undertaken, and when the maintenance inventory is connected to the PM's, inventory costs are also lowered by sourcing any parts required on a Just In Time (JIT) basis. Following completion of the service and updating the records on the CMMS a service history is created enabling management reports to be run evaluating issues such as time to next failure, and whether to repair or replace based on facts.

3 EXPENSE TRACKING

In a typical facility when an expense-tracking audit is undertaken it becomes clear very quickly that not all cost centres are equal when it comes to spending the maintenance dollar. In this exercise, it is appropriate to separate high cost asset repairs from general facility repairs. When comparing two cost centres engaged in similar work with similar activity levels it could be expected that repair expenditure rates should be similar. This is usually not the case. Some cost centres appear to have a disregard for costs generated relating to facility "repair" within their location. For example, the cost associated with an office reshuffle (moving furniture, repainting, rehanging notices boards, reorganising data and phone cables) each time there is a personnel change. Similarly, by identifying repair types, it becomes easier to identify staff behaviour pattens resulting in repair requests. Consider wall repairs, these can generally be related to two conditions: careless porters and / or insufficient wall buffering. Broken toilet seats in patient areas may be a result of clinical staff incorrectly transferring patients or a poor quality inventory item.

The ability to manage and track costs related to maintenance projects, particularly in times of maintenance restrictions due to budgets is crucial to performance management of cost centre managers. Disregard for the general repairs budget will have some cost centre managers resorting to understating the work required and job splitting (authorisation from different independent sources within the facility) to achieve small (and some times medium sized) projects. An example of an office space refurbishment to accommodate two clerks follows:

Authorised Works – 2 trades coordinated through maintenance approx $500 of work requested		**Works undertaken – 5 trades from various sources approx $1700 spent**	
Repaint	$350	Repaint	$350
Data point installation (1 computer port)	$150	Data Point installation	$150
Office desk (Relocate from existing stock)	$10	Refurbish existing desk	$125
Filing cabinet (Relocate from existing stock)	$10	Purpose built additional desk	$250
		Additional phone point installation	$150
		Additional phone	$35
		Refurbish and paint filling cabinets	$100
		Replace two office chairs	$250
		Purchase and install two notice boards	$75
		Installation of new ceiling lights	$125

It is crucial to track maintenance expenses generated by medical staff. Assets purchased by Doctors are generally not assessed under new product acquisition guidelines, which normally include an engineering assessment. This assessment includes: availability of local support and parts, reliability, PM costs. Consideration should also be given to loss of organisational revenue in down time waiting for parts. Thus creating a situation of potential under utilisation, (determinable through mechanisms such as Overall Equipment Effectiveness), and the loss of revenue blamed on the maintenance department because sourcing parts may require a 6-9 week lead-time. The requesting doctor changing hospitals or dropping the associated service before the life expiry of the asset of course compounds this situation. By combining mechanisms such as OEE, with a properly configured CMMS, an overall increase in asset availability of approx 5% is achievable.

On a broader scale in a maintenance department that is reactionary to breakdowns using non-controlled stores (or inventory), if KPI's such as Mean Time Between Failures and Mean Time To Repair are not regularly monitored and reported, a false indication of the state of the asset would be given by the fact of a high availability of the asset. The correct use of these indicators would tell us that we in fact might have a highly unreliable asset, which is costing an excessive amount of funds to keep the asset in service. Thus enabling the facility management team to make the repair or replace decision based on fact. A common mistake is not predetermining the level at which an asset goes beyond economic repair. A classic example is a syringe driver costing $1500, a typical breakdown will require a part worth $650 and $100 labour, therefore one repair will cost almost 50% of the initial purchase price generally at a point of 75% or greater of the life cycle of the asset. The asset in this case should be deemed as beyond economical repair.

Lack of inventory tracking is another significant cause of expense. Healthcare is not the only industry to suffer from untracked and unaccounted for inventory. The issues here are firstly the initial cost of the items (usually allocated to maintenance), which are then located about the facility, unaccounted for, and when consumed the cost is rarely applied to the cost centre consuming the item. The challenge here is to reduce the inventory to JIT items and clearly identified critical parts.. If for example you replace 10 fluorescent lights a week, it wouldn't make a great deal of sense to stock 200. The other issue here is standardisation wherever possible will also reduce the inventory. The next challenge is associating the cost of these items with a work order and allocating the total cost to the appropriate cost centre Clearly by controlling the inventory and attaching inventory to asset repairs a facility will not only gain a true cost but will not have thousands of dollars of inventory stock sitting on shelves. In most facilities, a saving of between 8-12% should be achievable.

The human resource element (the maintenance staff) is another area, which is not often under scrutiny to achieve set productivity levels within the normal working day. The questions to be asked cover two competing elements: first how many tasks are completed by each tradesperson on average each day. If this is evaluated with time taken for each, estimation can be made as to individual productivity; second, are tasks being undertaken just because staff are available. An analysis of job types and their priority in terms of mission criticality will indicated if over servicing is occurring or not. By moving to a planned or scheduled task approach there are further efficiencies to be gained over an unplanned or reactionary approach, it has been estimated by Daryl Mather (2006) that these efficiencies may be as much as 50% or greater.

In the instance where most elements of maintenance are outsourced, planning can deliver additional effectiveness for the maintenance department. For example the call out fee for outsourced trade could be at $75 per hour, if the task at hand is n urgent and only takes 15 – 30 minutes there is a productivity loss corresponding to the amount of time paid for (1 hour) but not consumed (30 – 45 minutes). By planning ahead and grouping non-urgent tasks into appropriate trade batches, and matching tasks to time units where able, the facility is able to reduce non-productive time across most repair orders raised.

However, the amount of savings actually realised will of course depend on the ability of the organisation to truly embrace the changes that can be achieved and to drive them through all levels of the company. It is not uncommon to see the full possible gains not realised in the first years of the implementation as the maintenance resources may, wisely, be deployed on reliability projects in order to usher in the next phase of maintenance development in a more controlled and permanent manner. Using this methodology, of reducing lost productivity hours from trades, we have achieved an overall reduction in maintenance expenditure of 5%.

4 TRIAGING REPAIR REQUESTS

The ability to cash flow a facilities standard maintenance repair costs is crucial to achieving budget while maintaining a functioning facility. Without a budgeted monthly cash flow for repair maintenance, it becomes exceedingly difficult for the maintenance manager to evaluate and correctly prioritise repair maintenance beyond all items presenting a safety issue. These must be removed from service and repaired as soon as practical, particularly if it is an asset critical for clinical service function. By moving work requisitions into a planned rather than reactive framework could provide savings against the maintenance budget.

To effectively triage the work requests the following management decisions need to be made:

- On a scale of 1(low) – 10 (critical) where does an asset sit
- On a scale of 1 – 10 the importance of the cost centre requesting work (e.g. Operating Theatre may rank 9 or 10)
- Grouped work categories – that is minor work of a similar nature

By evaluating a work request against these factors, it becomes easy to prioritise the list of competing demands into a planned work schedule. Correctly triaging, or prioritising work requests, the maintenance manager has the ability to group work types enabling a cost effective approach, minimising the number of call outs with associated non productive time. This is further developed by massaging grouped work requests into achievable (in terms of time/cost allocation) work units, defined as a block of work which fills a contracted unit of time, i.e. 3 tasks which amount to 1 hour of work, where the contracted unit of work is 1 hour). In this way, for example several requests for patch painting over a number of cost centres could be consolidated into one work order for an external contractor. This technique will also allow some work requests to be moved to another financial reporting period (week / month / year) as appropriate to satisfy cash flow or financial delegation requirements

5 MONITORING CONTRACT COMPLIANCE

The opportunity to capitalise over the life cycle of an asset does not rest solely with good management through the CMMS. While managing an asset through the CMMS will provide capital savings, if a facility is to maximise the potential for further saving it is essential that appropriately documented service contracts are established, detailing not only the asset service specifications but service contract KPI's. Another key issue is ensuring that asset maintenance has been undertaken in the manner prescribed to the specifications listed and in the periodic framework required. Mapping the actual services provided against the designated service specification, enables an analysis can be made as to the effect the servicing has had on the life and performance of the asset. This exercise will also illustrate service performance failures, deviations or non-compliances by the contractor. It will allow the astute maintenance manager to evaluate the quality of replacement parts used and their contribution to the asset life. An example of contractor failure in routine servicing of air handling units for a HVAC is the premature replacement of the drive belts twice in succession, the immediate questions are: has the contractor used the specified replacement part and did they check the pulley alignment at the first replacement? The answer was no in both instances when the correct part was installed and the alignment checked normal belt wear was experienced.

6 MAINTENANCE KEY PERFORMANCE INDICATORS (KPI'S)

To maximise the potential of an implemented CMMS it is essential that management defines the corporate direction and the business rules it applies to the facility corporate practices to enable the articulation of appropriate KPI's. It is also essential that KPI's be externally benchmarked where appropriate to like facilities and services to ensure that a best practice approach is adopted.

The approach I have used to determine KPI's was by categorisation of indication groups that would be required by management and then analysing what indicators would deliver a result. When comparing my list with that provided by Daryl Mather (2003) of Strategic-Advantages in his standardised approach to KPI's list, there was a high correlation between the two. The combined list of KPI's, from which you could expect to cover most the relevant issues pertaining to implementation of a CMMS, and produce data from which sound management decisions can be made, follows:

Definition of the high level KPI's used by the facility to monitor and control maintenance performance:

- **Unit costs for maintenance**
    - By Equipment
    - Per maintenance employee
    - As a measure against Estimated Replacement Value of assets
    - By cost centre
    - By Contractor
- **Overall Equipment Effectiveness**
    - Availability
    - Mean Time Between Failure (MTBF)
    - Mean Time To Restore (MTTR)
    - Utilisation
    - Quality
- **Defining what is capital works and what constitute operating costs.**
    - What criterion determines a plant improvement?
    - What constitutes like for like replacement?
    - Changes to Process and instrumentation drawings?
    - Changes to operating improvements?
- **Defining authorisation levels in dollar cost terms of each role within the maintenance organisation**. Thought needs to be given to bottlenecks that may be created from the facility delegations manual and the organisational structure.
- **Determining a prioritisation system that will allow the best use of resources across the facility**:
    - What resources to be used across various plants?
    - What resources to be used across various sites?
    - What are the structures of the work groups or teams to be included?
    - What work will be outsourced?
    - What are the time limitations to complete a task? e.g. max 30 minutes
- **Determining what are the definitions of various types of work orders.**
    - Safety
    - Capital
    - Preventive Maintenance – life cycle, and inspection / preventative
    - Repairs
- **Determination of the definitions of types of maintenance and setting levels to be used as benchmarks.**

  If there is a declaration by management on the internal benchmark level to be achieved, it needs to be clearly articulated that the benchmark levels <u>must</u> be derived from whatever maintenance strategy is adopted by the facility. External benchmarks in this instance do not necessarily apply to your facility unless there is a similar match up in the strategic approach and from a “like” industry or facility.

  - Preventative Maintenance – %
  - Predictive Maintenance – %
  - Corrective Actions – %
  - Breakdowns - < %

- **Standing work orders** – Their needs to be an initial focus on this area, determining which items need to be covered by standing work orders. Although there are many approaches, it is best to utilise these for overhead items such as training, holidays and breaks. Use of these as blanket work orders elsewhere will blur the results available from the CMMS system.
  - **Defining business processes and the KPI measurements required for controlling these.**
    - Backlog management
    - Age by Priority measurement
    - Number of safety work orders (trended)
    - Planned work orders per work group
    - Planning/ Scheduling systems
    - Planned/ Scheduled ratios, again these should be industry specific in terms of tolerances, and a risk analysis should be undertaken as part of the process of determination.
      - Planned / Scheduled – %
      - Planned / Unscheduled – %
      - Unplanned/Unscheduled - < %
      - Stores Service rates - %
      - Work completed on time by priority schedule
      - Execution and Data capture systems
      - Highlight the standard of coding required to provide a base for future analysis
      - Fault causes?
      - Duration between Work Order and Work Completion
      - Parts used?
      - Standardised text entries for free text?
- **Engineering works**
  - Criteria for review and execution stages
  - Monitoring / measurement of effectiveness
  - Integration for execution
  - Criterion for inclusion, including justifications process
- **Analysis and actioning loops**
  - Root Cause Analysis processes and reporting required for accurate targeting of the "critical few" items.
  - Measurement of effectiveness

7 IMPLEMENTATION, TRAINING AND EMBEDDING

To succeed, a maintenance program must be embedded into the culture of both the maintenance department and the facility and delivered with role specific training to staff throughout the facility and contractors.

Introducing the concepts raised in this paper to a facility must be undertaken within a change management framework. Initially, it is imperative to have a vision of what you are trying to achieve. The David Firth (1999) has identified the Seven Factors needed to implement Vision as:

- The vision is freely chosen
- Consideration of alternatives has been made
- Clear knowledge of consequences
- Prized and cherished by you
- Publicly proclaimed throughout the organisation
- It is acted upon (repeatedly), not just talked about.

"If your intentions are clear, you create an electric field of possibility that actually pulls creativity out of yourself and those around you. This is the real power of intention. It inspires you in ways you could not predict… Growth is an excitement to individuals and companies and comes from stretching to achieve things that may not have seemed possible even a week earlier" Gay Hendricks: The Corporate Mystic (cited in Firth, 1999)

When you have established your vision for the maintenance program, it is important to understand the dimensions of the Corporate Culture within the facility. This entails understanding the current cultural climate within which the facility operates, considering the organisations business focus and how the leadership and management functions. It is also important to understand the structure of the organisation and the resources it has available to undertake your project, which should include personnel development and customer focus. Principles, which should be applied for success in building consensus for change, are: Involvement, Education, Honesty, and Congruency. Further discussion on the change management required for implementation has not been included in this paper due to word limitations.

Training will need to be developed, or adapted, in using the system to accommodate the facility business rules and processes of the facility. The focus of the training is such that each of the employees knows what their responsibilities, accountabilities and what the role relationships with others in the organisational structure are. Elements that should be considered when estimating the training required not only include the maintenance staff and various department heads around the facility, it should also include the contractors. Here it is important to ensure that the relationship that is developed with contractors enables them to function effectively within your facilities environment. This process commences with a fully documented contractor induction, complete with a contractors handbook, will commence the integration into your facilities culture. It is important that contractors understand where your organisation is coming from and what the organisation needs from them, both in service delivery and how they conduct their business on site. To facilitate improved communication between contractors and your internal customers an important consideration is the introduction of contractors to key staff members. Remember that each department will have it's own operational issues in regards to how a contractor can perform a service within that area.

8 REFERENCES

[1] Berger, D. (2003) Asset Management – controlling the life cycle. http://www.plantservices.com/voices/asset_manager.html

[2] Mather, D. (2006) Calculating the Savings from Implementation of CMMS.

[3] http://www.maintenanceworld.com/Articles/matherd/Calculating-Savings-Implementation-CMMS.htm

[4] Mather, D. (2003) Developing CMMS Implementation Templates http://www.maintenanceworld.com/Articles/matherd/developingcmms.html

[5] Firth, D. (1999) Smart things to know about change. Oxford, UK. Capstone Publishing Limited.

Acknowledgments

Possumworks Management Consultancy

http:/www.3pi.com.au

http://www.ferret.com.au

http://www.maintenanceworld.com

http://www.plantservices.com

Gay Hendricks: The Corporate Mystic

FIELD INVESTIGATION OF CAVITATION AND FLOW INDUCED VIBRATIONS IN SUBMERGED VERTICAL PUMPS IN A POWER PLANT

Tan Chek Zin[a], M. Salman Leong[a] and Lee Yuk Choi[b]

[a] *Institute of Noise and Vibration, Universiti Teknologi Malaysia, Jalan Semarak, 54100 Kuala Lumpur, Malaysia.*

[b] *Faculty of Mechanical Engineering, Kolej Universiti Teknikal Kebangsaan Malaysia (KUTKM), Locked Bag 1200, Ayer Keroh, Melaka, Malaysia.*

Abstract: Pumping sets are common but often critical operating assets in the industry. Such pumps often do not have adequate pressure and flow instrumentation that could detect abnormal hydraulic operating conditions. This paper presents field investigations of two submerged vertical pumps with suspected cavitation and flow induced vibrations in a power plant. These pumps had a history of catastrophic impeller failures. The vibration analysis allowed an assessment to be made on the pump of concern; upon which visual examination of the newly installed impeller was made that confirmed the occurrence of cavitation. High frequency random vibration was shown to be a good indicator for flow excitation and cavitation. Comparison of results under service conditions between a cavitating and non-cavitating pump are presented. Failure modes showed cavitation erosion and a sheared impeller blade. Evidence of excessive clearance and/or sleeve bearing wear was also noted in the pump of concern. The investigations confirmed concern relating to operations during low tide and combined pumps operated in parallel where inadequate submergence was identified as a likely cause to the impeller failures. The investigations implied a system design problem and pump operating conditions compounding the system design error that had resulted in frequent failures of the pumps.

Key Words: Cavitation, Flow excited vibrations, Pump, Detection, Diagnosis

1 INTRODUCTION

Pumping sets are common and often critical operating equipment in the industry and in particular in the power generation and petrochemical industry. This paper presents field investigations of cavitation and flow induced vibrations in two submerged vertical pumps (referred herein as SVP1 and SVP2) used for pumping of cooling water to the condenser of a steam turbine in a power station of an Independent Power Producer (IPP) in Malaysia. The Station has 2 unit gas turbines with an additional unit steam turbine, operating as a combined cycle plant with steam generation from a waste heat recovery boiler. The rated capacities of the gas turbines are 2 x 110 MW, and an additional 110 MW from the steam turbine; with a total plant power generation output of 330 MW. The IPP commercial obligations require total load dispatch from the Station being a base load plant to be operated in a combined cycle mode. In the event that the steam turbine is not operational, any electricity generation from the gas turbine is not accepted for dispatch to the national electricity grid. Failure of the submerged vertical pumps would potentially result in a total plant shutdown. Criticality of these submerged vertical pumps is obviously very high.

The pumps were identical (make, model and operating speeds) and installed in parallel connected to a common header after discharge routed to the condenser. The pumping medium is sea water. Seawater is drawn from a suction pit connected to the sea via a water intake tunnel. A photograph of the pumps is given in Figure 1(a) and a schematic drawing in Figure 1(b). While the pumps could be operated either independently or in combined pumping modes - operating continuously 24 hours per day, both pumps need to be operating when the turbines are operating close to base load (full load capacity) conditions. In the event that only one pump is operating, some load restriction would be required as there would be insufficient cooling water to the condenser. Each pump has 4 blades, and the pump vertical shaft supported on cutless rubber bearings (sleeve bearings) located in an inner tube in the riser tube. The motor had rolling element bearings. The drive shaft is approximately 10.5 meters from the sea level. The design specifications of the pump are total dynamic head (TDH) of 16.2m, flow capacity of 13,800 m^3/hour, and operational speed at 595 rpm. The pumps are axial flow with a specific speed (Ns) in the order of 8500. The manufacturer's recommendations for minor inspections (maintenance) are scheduled on an annual basis, and major overhaul scheduled once every 5 years.

2 NATURE OF PROBLEM AND MANAGEMENT ISSUES

Both pumps had a history of premature impeller failures, and had been an issue of concern to the plant with respect to unscheduled breakdowns and interruptions to electricity load dispatch (noting that there would be some reduction in plant capacity whenever a pump becomes unavailable, and a total plant shutdown when both pumps are not available). While it was originally thought that the problems associated with these pumps were a mere case of unreliable machinery and thereby having to fix the pump whenever a failure occurs; the work and sequence of trial and error remedies to have the pumps brought back to service points to a potential system design error and operational conditions compounding the error. There are currently no flow measurement devices on the pumps, and came fitted with an analogue pressure indicator on the common discharge pipe only. As such it was not possible to verify actual pump operating conditions on a single pump basis or on a combined system operating mode basis. The plant management obviously requires identification on the root cause(s) to the pump failures, and whether there is indeed a system design problem with the pumping system such that a long term solution for improved pump reliability could be identified. In the event of a system design error significant remedial works shall be necessary and in all likelihood can only be implemented during a long major outage, preferably to be scheduled with one already planned in two years time (2008). A brief history of the failures and attempted solutions of fixing the problem is described below. A table summarizing key observations and the corresponding site measurements for vibrations is given in Table 1.

Table 1: Summary of recent failures in the pumps, work done to the impeller (Pump SVP1) and vibration investigations

Date	2002 - 2005	1 October 2005 (1st Visit)	18 Jan 2006 (2nd Visit)	25 March 2006 (3rd Visit)	26 May 2006 (4th Visit)
Pump SVP1	Original impeller. Catastrophic failure of impeller in Dec 2003. Broken blade welded back in Dec 2003. Failures to impeller in Sept 2004 and Nov 2004	New impeller of different design (different shape compared to original impeller)	Refurbished (original) impeller	Refurbished (original) impeller	New impeller (original design)
Pump SVP2	New pump replacement in 2002 (with original impeller design)				
Vibration Site Measurements	No vibration history available.	Vibration measurements undertaken with SVP1 and SVP2 in combined operating mode	Vibration measurements undertaken with SVP1 and SVP2 in single operating mode	Vibration measurements undertaken with SVP1 and SVP2 in combined operating mode	Vibration measurements undertaken with SVP1 and SVP2 in combined operating mode
Condition of SVP1 impeller	Frequent failures of the impeller.	New impeller (different design) was removed at end Oct 2005 and replaced with refurbished impeller of original design		Refurbished impeller failed on 21 April 2006, and replaced with new impeller (original design).	

These pumps were installed in 1997 when the power plant was converted from an open cycle to a combined cycle plant. While there had been problems with the pumps, there was however no records for these pumps to assist plant personnel currently responsible for its maintenance to historically track the failures since installation. There was however a total pump replacement (herein referred as SVP2) in 2002 that was necessary as a result of a catastrophic failure of the original pump arising from a broken impeller. Ever since this pump replacement, minor pitting on the pump blades of SVP2 was noted during periodic maintenance and inspections. While minor cavitation was suspected in SVP2 based on the visual evidence from the pitting of the blades, the unit had however been operating relatively trouble free since 2002. The other pump SVP1 is however more problematic. There had been unscheduled breakdown failures (of more than 10 times over the past three years), with a major catastrophic failure due to a broken impeller during end December 2003, Figure 12. Instead of an entire pump replacement in this case (as compared to SVP2 in the previous year) the broken impeller of SVP1 was welded back on January 2004, and a dynamic balancing undertaken prior to its return to service. After 9 months of operation a similar failure occurred on the impeller. Another welding job was undertaken to return the impeller back into service in September 2004. After two months of operation, the impeller failed again for the third time in similar pattern.

This pump was then rendered unavailable for a relatively long period of time as attempts were made to seek a resolution to the problem. The pump vendor then recommended a new impeller with a revised design (with a reduced trailing edge profile perhaps in an apparent attempt to reduce forcing on the impeller blades). The pump with the revised impeller was then installed in September 2005. Almost immediately upon operations of the pump (SVP1) with the new revised design impeller, cavitation was suspected since vibration levels were excessively high with a corresponding rattling sound from the pump clearly audible. The absence of OEM fitted sensors for pressures and flow meant that pump operating conditions could not be monitored for any onset or occurrence of cavitation. There were OEM's fitted vibration sensors on each pump, but with readings limited to overall amplitudes only. SVP1 had vibration levels of 9 mm/s (as noted on 1st October 2005), whereas vibrations of the other relatively trouble free SVP2 were typically in the order of 3 mm/s. The new impeller of revised design as fitted to SVP1 was immediately removed, and a refurbished (original) impeller was fitted back in October 2005. The refurbished impeller failed (yet again) in April 2006. A totally new impeller replacement with the original design was fitted in May 2006.

The above behavior of SVP1, while in hind sight should have been expected, nevertheless posed a dilemma to the plant personnel and senior management. Frequent making "good" a broken impeller, failed attempt in a new impeller design (albeit on the recommendations of the pump vendor/OEM) and a new impeller replacement (of the original design), and not forgetting a new pump replacement for the other pump SVP2, suggested the inability to identify and remedy root cause(s) of the problems. Continuing operations in such a manner cannot warrant trouble free operations, improved pumps reliability and plant availability. Vibration measurements were used to help in the potential identification of the root causes correlated with other visual observations under a situation of inadequate instrumentations for actual pumps performance verification. The vibration based investigations presented herein strongly implied a system design problem and pump operating conditions compounding the system design error.

3 CAVITATION AND ITS DETECTION

Physical examination of the impellers when removed for inspection showed evidence of cavitation for all impellers. The preferred method to determine cavitation is by pressure measurements [6-9]. In the above pumps, there were no suction pressure indicators for such measurements in particular. As the pumping sets did not have any instrumentation for flow and pressure measurements to confirm occurrence of cavitation, vibration analysis was initiated in an attempt to confirm the onset and/or occurrence of a cavitation. This also presented an opportunity to examine the feasibility and the practicability of the detection of cavitation as a field validation of an experimental study of the authors [5].

Cavitation usually occurs in a flowing liquid where the local pressure drops below the saturation vapor pressure of the liquid at a given local temperature, causing liquid evaporation and generation of vapor bubbles in the low pressure region. Fluid machinery is a common application where low pressures are routinely generated by hydraulic action, for example on blade surfaces, with a consequent possibility of cavitation. The existence of cavitation is undesirable as it results in deterioration of the hydraulic performance, lead to physical damage affecting the equipment or component structural integrity and higher noise. Onset of cavitation is defined by ISO 3555 as a 3% drop in the pump pressure head without inlet or discharge throttling at a given flow rate. A characteristic audible sound (rumbling stones or 'marble' sounds) would be emitted from the pump when cavitation is occurring.

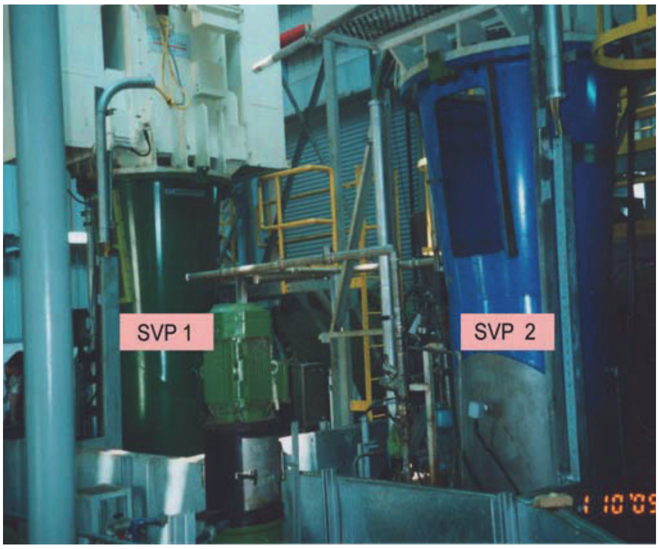
Detection of cavitation in pumps has been by subject of interest. Jeremy Jensen (2000) reported that the dynamic pressure was a direct indicator of cavitation, whereas Net Positive Suction Head Available (NPSHA) monitoring was an indirect indicator. Jorge L. Parrondo (1998) reported that pressure at outlet pump is sensitive to changes when cavitation occurs. The sensibility of pressure at the inlet of the pump was also regarded as a satisfactory indicator. Experiments have shown that there is a discrete frequency tone with the audible noise spectra, typically at half of blade passing frequency (BPF), which is strongly dependent on the cavitation process and its development (M.Cudina,2002). The discrete frequency tone at half of BPF could therefore be used to detect the incipient of cavitation and its development. T. Uchiyama (1998) investigated a fault diagnostic method for centrifugal pump using neutral network approach.

With the use of computational fluid dynamics (CFD) to analyze the viscous, turbulent flows, vortices in fluid machinery in presence of cavitation could be visualized in a computational model. These tools are however fairly complex and not readily available to the plant personnel in practice. A more conventional and robust method for the detection of cavitation and flow induced vibrations in general based on vibration measurements represent the most viable tool to the plant personnel.

4 FIELD MEASUREMENTS

Vibration measurements were undertaken on both pumps with different operating modes and impeller fitted for pump SVP1. Table 1 tabulates the respective vibration measurements undertaken for corresponding operating conditions. Vibration spectra and time waveforms of the respective pumps were measured and assessed. In the more recent site surveys bump tests were also undertaken on the pump casing and piping when the pumps were not operating to identify structural resonances For the purpose of this paper only selected results are shown and discussed for brevity. Since prior baseline measurements were not available, a comparison between the two pumps were undertaken with the intent of comparing vibration characteristics of the pump of concern (SVP1) with the pump that was regarded to be in fairly good conditions (pump SVP2 on the basis that this pump had been operating trouble free since its replacement in 2002). This was deemed a reasonable approach since both pumps were identical (same model and same operating condition) operating at a similar running speed (585 rpm). There may be some differences in the structural mobility due to its physical installation next to one another although measurements were undertaken at almost similar positions on the respective pump sets. Measurements and comparison of the measured data for SVP1 and SVP2 were undertaken in similar location and directions.

Measurement at the commonly preferred location at the bearing was not feasible as there was no convenient access to the bearing. Measurements were therefore restricted to secondary locations on the pump casing. Additional measurements were undertaken on the pump discharge base flange, Figure 2(a), at the mechanical seals, Figure 2(b), and at the motor.



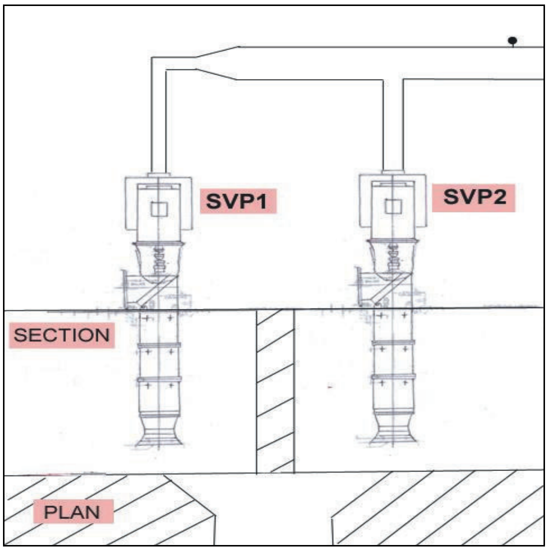


Figure 1(a): SVP1 (in green color) & SVP 2 (in blue color) located adjacent each other

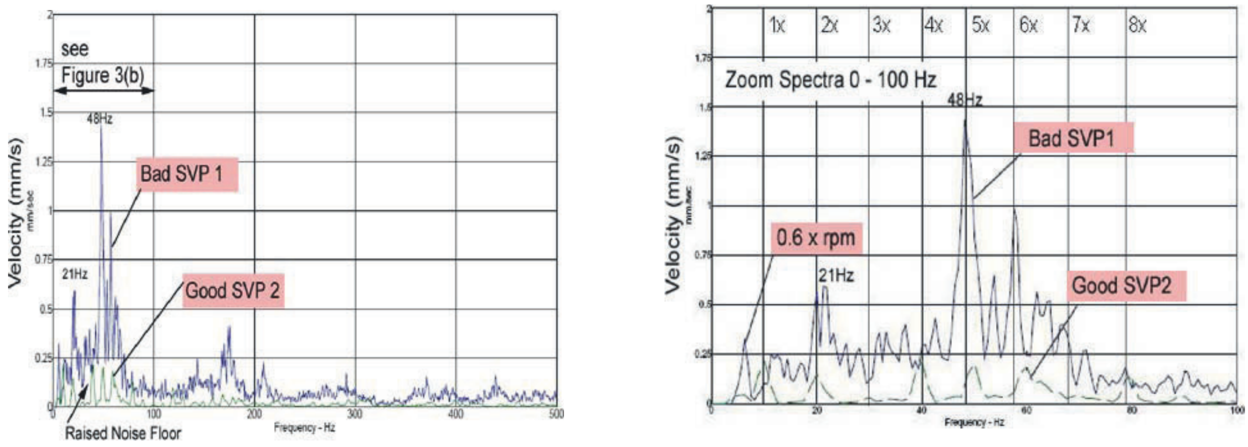
Figure 1(b): Schematic drawing of pumping installation

Figure 2(a): Accelerometer mounted on pump discharge base and casing

Figure 2(b): Accelerometer mounted on mechanical seal

5 RESULTS

A series comparison between the vibration data was taken on various locations from SVP1 (fitted with the new revised impeller design) and SVP2 are shown in Figure 3 to Figure 8. The blue line (solid line) as shown in the following figure presents vibration data for SVP1, while the green line (dashed line) presents vibration data for SVP2. Results comparing the original impeller (blue line) and new revised impeller design (red line) for the same pump SVP1 are given in Figure 9 to Figure 11.



(b)

Figure 3: Comparison of casing vibration of SVP 1 (revised impeller) & SVP 2 (Casing pump discharge - Horizontal)

Vibration velocity spectra measured on pump discharge casing are given in Figure 3(a), and the zoom spectra in Figure 3(b). The spectra for the "bad" pump SVP1 showed pertinent vibration response peaks at 6.25 Hz (~0.6x rpm), 21 Hz (2x rpm), 31 Hz (~3x rpm) 48Hz (5x rpm), and 58 Hz (6x rpm); Figure 3(b). The blade passing frequency at 4x rpm and 2x BPF (8x rpm) was not so evident in this bad pump. Multiples of harmonics of up to 8x rpm and raised noise floor in the spectrum of the bad pump SVP1 were therefore clearly evident. This suggested that clearance related problems where sleeve bearings wear to be present. The "good" pump SVP2 did not demonstrate high vibration response peaks, nor was the harmonics multiples so clearly evident. The blade passing frequency at 4x rpm and 2x BPF (8x rpm), although of lower amplitudes as compared to the bad pump, was nevertheless present in the good pump SVP2.

The dominant vibration response peaks of 48 Hz while suspected as a possible resonant peak was however not confirmed from the measurements. A subsequent bump test undertaken on the pump discharge casing showed a resonant frequency at 21 Hz for the pump casing, and a resonant frequency of 31 Hz for the inner tube of the riser tube (holding the shaft).

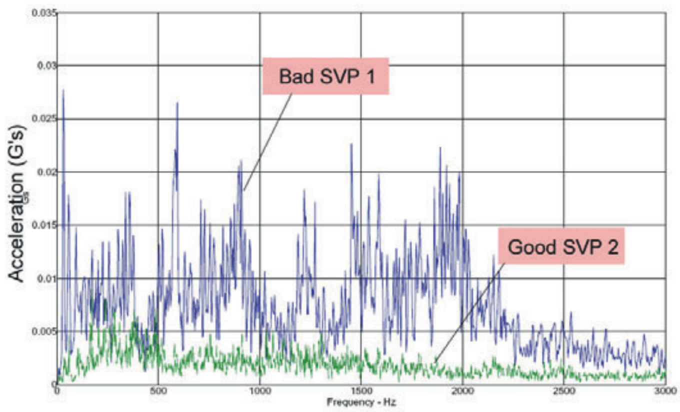
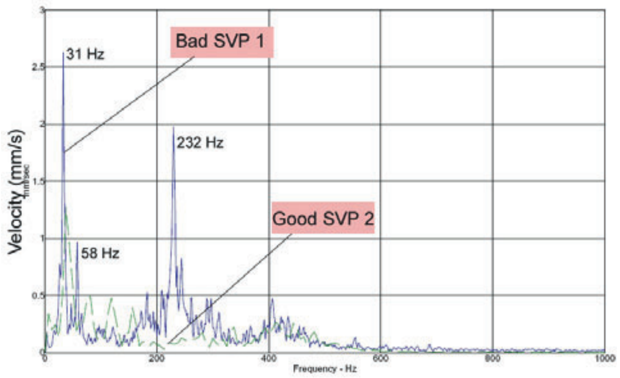


Figure 4: Comparison of casing vibration of SVP 1 (revised impeller) & SVP 2 (Pump discharge base - Axial)

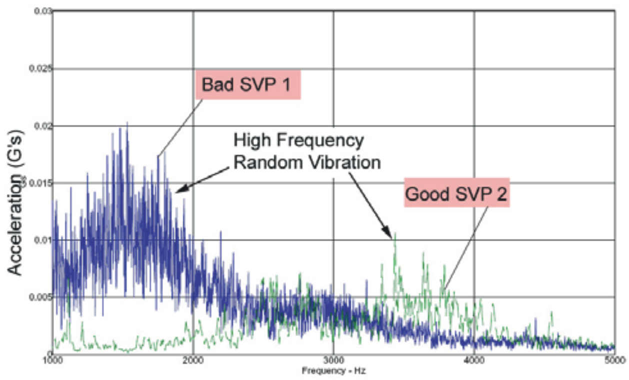
At the pump discharge base (measured in axial direction), vibration acceleration amplitudes of the bad pump SVP 1 were significantly higher than the good pump SVP2 (Figure 4); with overall vibration amplitude of 0.20 Gs for SVP1 and 0.07 G's for SVP2. Impulsive vibration peaks (spikes) and high frequency random vibrations were very evident in SVP1. Random vibrations at high frequency are often associated and recognized as characteristics for cavitation and flow turbulence (flow excited vibrations). The vibration spectra peaks of the good SVP2 as compared to SVP1 were significantly lower. The measurements therefore confirmed presence of cavitation and flow induced vibrations.

Measurements at the mechanical seals in the axial direction showed increased vibration velocity response peaks at 31Hz (~3x rpm), 58 Hz (6x rpm), 232 Hz (6x BPF), Figure 5(a). The resonant frequency of 31 Hz for the inner tube within the riser tube as excited by the flow was particularly evident. The 232 Hz peak (although at 6x BFP) currently remains unconfirmed as to whether this is a component resonant frequency. Random vibrations were evident in the frequency range from 150 Hz to 500 Hz in the acceleration FFT for both SVP; but were more pronounced and significantly higher in the bad pump SVP1, Figure

5(a). In the bad pump SVP1, Figure 5(b), the random vibration tended to be in the frequency range from 1000 Hz to 4000 ; and in the good pump the random vibrations, although less intense, were from 2000 Hz to 4000 Hz.



(a) Velocity spectrum (0 – 1000 Hz)



(b) Acceleration spectrum (1000 – 5000 Hz)

Figure 5: Comparison of vibration of SVP1 (revised impeller) and SVP2 (Mechanical seal – Axial direction)

Tri-axial vibration measurements were also undertaken on the motor. Comparisons of the vibration response are given in Figure 6. It was confirmed that vibrations of SVP1 were more severe than SVP2 in all measured directions (axial, horizontal and transverse). The axial vibration velocity dominating peak response of 23.1 mm/s was at 29.5 Hz (3x rpm), with other pertinent response at 6.25 Hz (0.6xrpm) 1.4mm/s and 10 Hz (1x rpm) 1.0 mm/s. Visual observations confirmed severe shaking of the SVP1 motor at the top. The 30 Hz response (as marked up in Figure 6) was identified to be a resonant frequency of the inner tube within the riser tube to hold the shaft; and in this instance was most likely to be excited by random vibrations associated with flow excitation present in the bad pump SVP1 (as evident from Figure 5 and Figure 6 above).

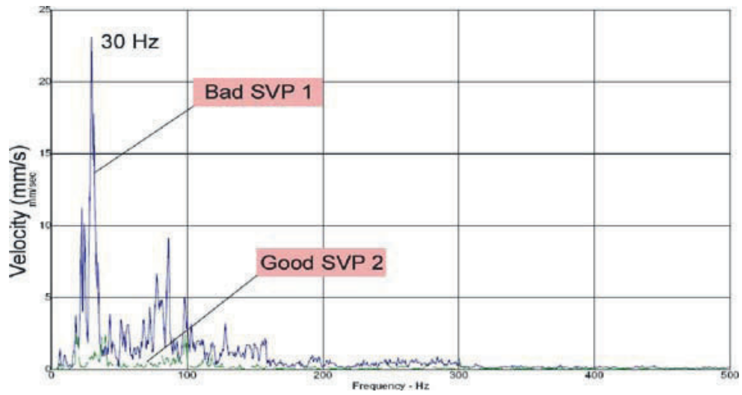


Figure 6: Comparison of vibration of SVP1 (revised impeller) and SVP2 (Motor casing - axial direction)

In the horizontal direction, vibration at SVP 1 was marginally higher than SVP2 at frequency of 50 Hz to 300 Hz, Figure 7. In the transverse direction vibration response peaks of 6mm/s vibration at 6.25 Hz, 5.7mm/s at 45 Hz, 4.3 mm/s at 57 Hz, and 5.9mm/s at 83 Hz were noted, Figure 8. The vibration amplitudes showed a more severe vibration in the transverse direction than the horizontal direction for the bad pump SVP1; in all likelihood due to the flow direction of the sea water towards the bell mouth of the submersible pumps in the transverse direction.

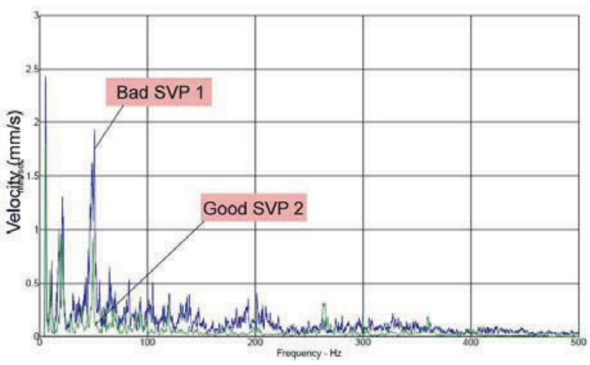


Figure 7: Comparison of vibration of SVP1 (revised Impeller) and SVP2 (Motor horizontal direction)

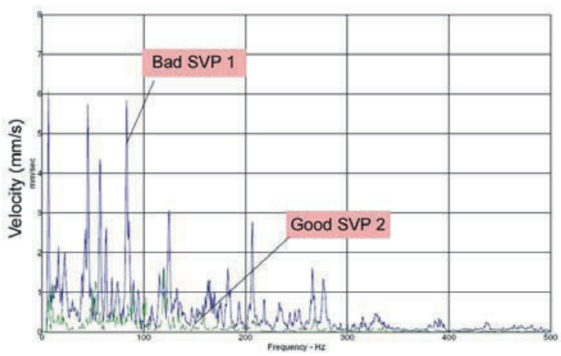


Figure 8: Comparison of vibration of SVP1 (revised impeller) and SVP2 (Motor transverse direction)

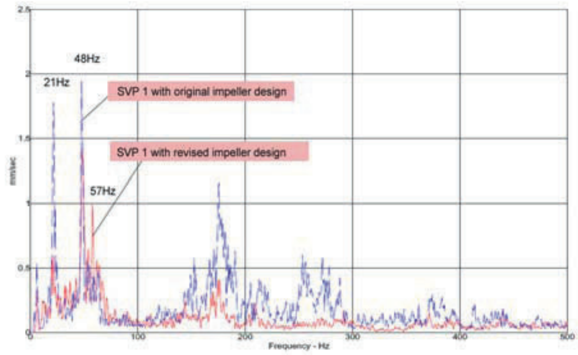


Figure 9: Comparison of casing vibration of SVP1 with original and revised impeller

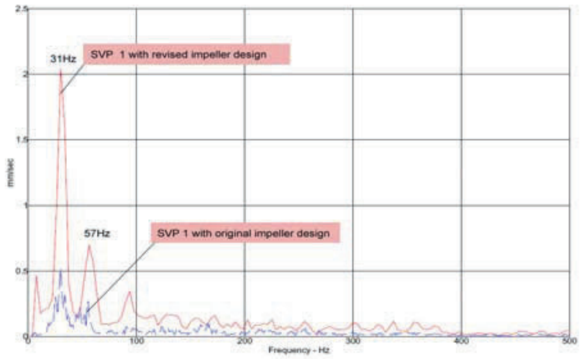


Figure 10: Comparison of pump discharge base vibration of SVP1 with original and revised impeller

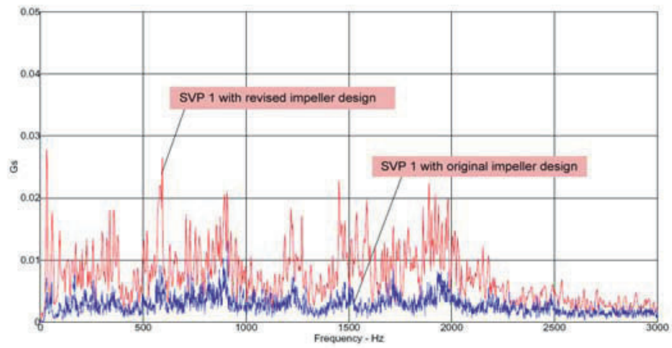


Figure 11: Comparison of high frequency vibration spectrum of SVP1 with original and revised impeller

The above figures (Figure 9 to Figure 11) present typical comparisons of measured vibration response comparing the original impeller and new revised impeller design in the same pump (bad pump SVP1). The measurements showed higher vibration response with the original impeller design, where the resonant peak of 21 Hz and 48 Hz associated with resonant frequency of the casing (21 Hz) were prominent, Figure 9. This suggested higher flow excitation to the casing with the original impeller design. Measurements at the pump discharge base, Figure 10, and at the mechanical seal however showed a lower vibration response with the original impeller design. The structural resonant frequency peak of 31 Hz of the inner tube within the riser tube was particularly evident. Random vibrations associated with flow excitation and turbulence was also higher in the pump with the revised design impeller, Figure 11. This was consistent with the relatively more severe cavitation in the pump with the revised design impeller.

6 DISCUSSIONS

High frequency random vibrations commonly associated with flow turbulence and cavitation were very much evident in the bad pump SVP1, particularly when viewed (i.e. measured) in the acceleration parameter. This was particularly evident for measurements undertaken at the pump discharge base and the mechanical seals. In the good pump SVP2, high frequency random vibration could also be observed on the mechanical seal although generally less pronounced (severe). Some nominal random vibration was also observed for the pump discharge base at this good pump SVP2. Structural resonances of the casing and components making up the riser tubes were excited by these random vibrations. The site measurements confirmed acceptable vibration signals for measurement undertaken at locations of the pump discharge base and mechanical seals, in the absence of permanent vibration sensors installed at the submerged bearing housings. Signal strengths and data that could be extracted from measurements at these secondary locations were consistent with observations in experimental studies and recommendations in the literature. Visual inspections of the impeller and blades confirmed occurrence of pitting associated with initial stages of cavitation and/or combination of erosion.

Vibration severity in the bad pump SVP1 was consistently higher than SVP2 at all measured locations. The vibration velocity spectra of the bad pump SVP1 also suggested excessive clearance and/or wear in the sleeve bearings which could be a consequence of the excessive vibrations in the bad pump SVP1 (and less so as a root cause of the severe vibrations).

The above measurements and aural evidence from the pumps confirmed the need for the pump of concern to be opened up for visual examination on several occasions. The rotor shaft and impeller assembly when removed from its casing and the riser tube in several instances confirmed a damaged impeller. A photograph of the damaged impeller in one of the several cases of impeller failure is shown in Figure 12. The failure mode appeared to be a blade shearing off at the root connection with the central hub (where stress concentration is likely to be highest). Such a failure is more consistent with excessive hydraulic forcing than erosion wear associated with cavitation. In the case with the revised impeller design, cavitation was evident from the pitting of the impeller blades which occurred on the newly installed impeller just after 3 weeks of operation, Figure 13 and Figure 14.

The measurements and failure history suggested that the second pump SVP2, while being identical to the bad pump SVP1, has a less severe vibration problem with less severe hydraulic excitation although showing evidence of some cavitation. It is therefore possible that there are differences in the hydraulic operating point of pump SVP2 which resulted in a less severe state of cavitation and flow induced vibrations. Differences in water entry conditions at the bell mouth and material differences in the impellers between the two pumps (implying that the impeller of SVP2 had a higher degree of resistance to the erosion and pitting from the cavitation) may also be possible. Not withstanding the above observations, it ought to be recalled that this pump SVP2 did suffer earlier failures that required a complete pump replacement. Audible 'rattling' noise of the bad pump SVP1 was particularly more pronounced, and less obvious in the good pump SVP2. This correlates well with the relative severity of the vibration amplitudes and vibration spectra patterns of the two pumps.

Figure 12: Damaged original impeller of the SVP1 (note displaced position of damaged blade)

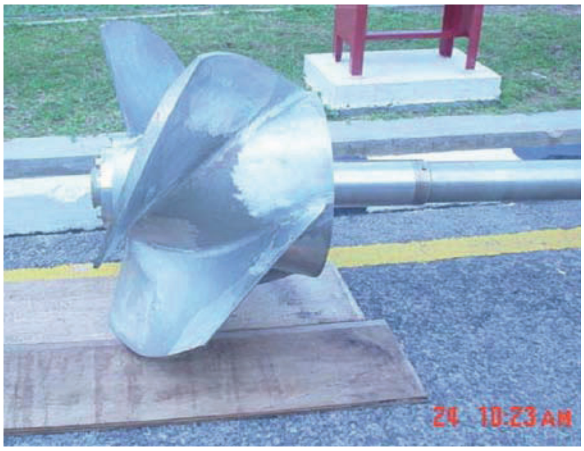
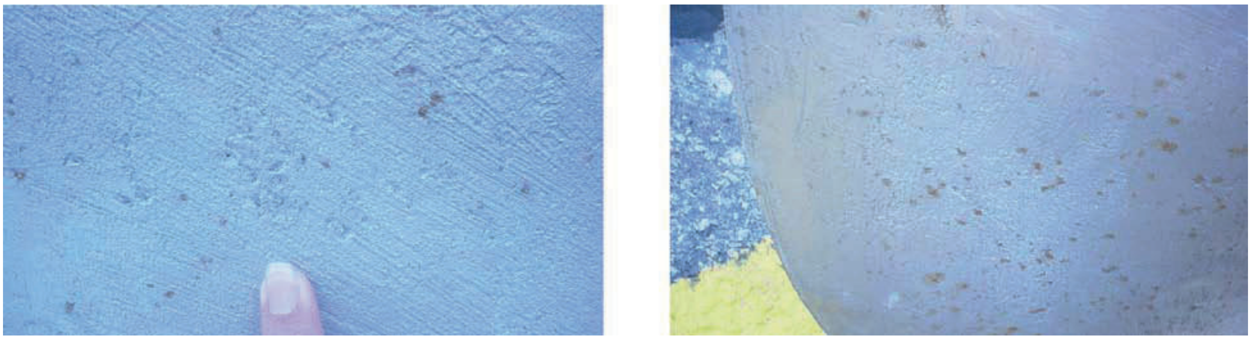


Figure 13: New impeller (revised design) of the SVP1 after 3 weeks operation



(a) **(b)**

Figure 14: Pitting found on the blades after 3 weeks operations for new (revised design) impeller.

While cavitation was more evident in the bad pump SVP1 than the good pump SVP2, it cannot totally explain the recurring failure mode of a sheared blade at the root of the impeller, Figure 12. This therefore suggested the presence of excessive hydraulic forces arising from flow turbulence in addition to cavitation in the pump. Resolution of the flow induced vibrations (flow turbulence) and cavitation problems would require a review of the pumps system design, and in particular the suction heads of these submersible pumps and the sump pit design. During high tide of the sea, the submergence of the pumps was approximately 5m from the bottom of the sump. At low tide, the water level would drop to between 3 to 4m from the bottom of the sump; Figure 15. When both pumps were operated simultaneously in a combined mode, the water level at the sump pit dropped further by almost 1m as compared to the case when only one pump was running. This inevitably resulted in a situation when the pumps are now at/or below a critical operating zone below the minimum submergence required by the pump. Inadequate submergence (i.e. a drop in the suction head) would potentially result in a cavitating state. This would be consistent with vibration measurements that showed acceptable vibration levels when only one of the two pumps is operating as compared to high vibrations when both pumps are operated in a combined mode. Visual examination of the water surface in the sump pit during pumping operations nevertheless showed steady (still) water surface without undue turbulences or vortices.

The fact that there is a drop in water level in the sump pit when the pumps were operated in a combined mode (full capacity) suggested that there may be a system design error with the sump pit where sea water entry at the water intake tunnel may be undersized (or obstructed). The water level also drops to a situation with little or no margin above the minimum submergence level during low tides. As such cavitation and turbulence within the impeller may be inevitable.

In axial flow pumps, cavitation and flow turbulence cannot be explained in the same way as radial-flow and mixed flow pumps. The water enters an axial-flow pump in a large bell-mouth inlet and is guided to the smallest section (called the throat) immediately ahead of the impeller. The capacity at this point should be sufficient to fill the ports between the impeller blades (assuming normal full submergence). When the head is increased beyond a safe limit, the capacity is reduced to a quantity insufficient to fill up the space between the impeller, creating cavities of almost a perfect vacuum. When these cavities collapse the water hits the impeller with a force sufficient to pit the surface of the vane. Conventional good practice to avoid cavitation in axial flow pumps suggests avoidance of the following:

- i) Heads much higher than the head at peak efficiency of pump;
- ii) Capacity much lower than the capacity at peak efficiency of pump;
- iii) Suction lift higher or submergence head lower than manufacturer's recommended levels;
- iv) Heads much lower than the head at peak efficiency of pump; and
- v) Capacity much higher than the capacity at peak efficiency of pump.

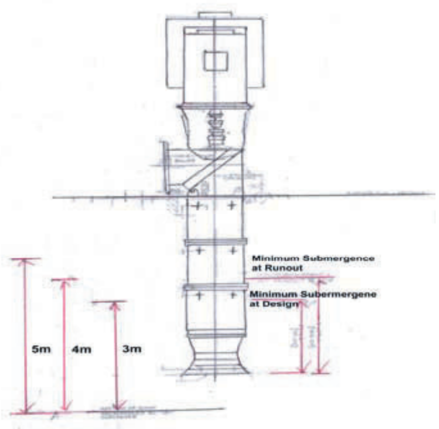


Figure 15: Minimum submergence required

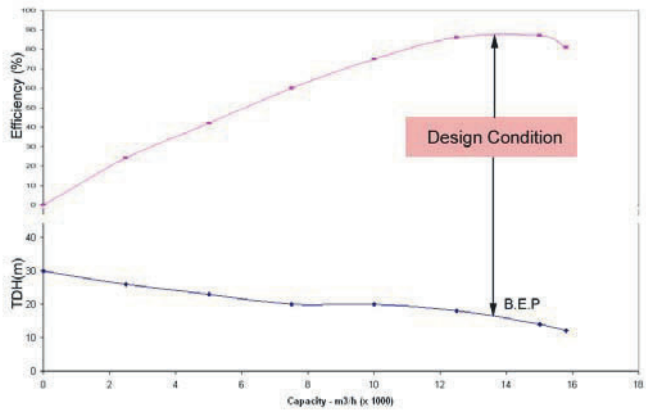


Figure 16: Design performance pump curves for SVP 1 & SVP 2.

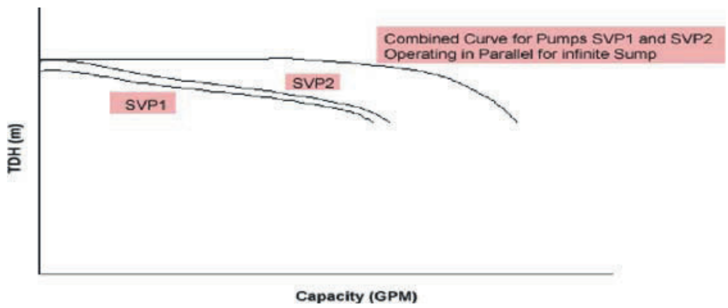


Figure 17: Typical performance curves for individual and pumps operating in parallel

As the pumps SVP1 and SVP2 do not have proper pressure and flow rate measurements devices or indicators, it was obvious that an immediate task is to install pressure measurements sensors on the suction and delivery sides. A flow measurement device (either by direct measurements or indirect measurement devices such as a Yates meter) should also be considered to confirm the pump operating points. Consistent with good practice, the operating point should be at its best efficiency point (B.E.P). Cavitation is generally not expected to occur if the pump is operated in close proximity to its B.E.P., Figure 16. In the absence of any pressure or flow measurement instrumentations (except for a simple analogue indicator on the common header of the discharge pipes) in the current situation, it was not possible to determine actual operating points of the pumps on an individual basis and when operating parallel in a combined mode. As it stands the system performance curve in a combined mode is not unknown (other than what is shown in Figure 17 as a hypothetical curve for an idealized infinite sump), nor are the actual operating points of the respective pumps in a combined mode (and individual mode) known. It is also uncertain if the pumps are operating at the same hydraulic state (balanced state) when operating in a combined mode.

7 CONCLUDING REMARKS

The investigations strongly implied a system design problem and pump operating conditions compounding the system design error that had resulted in frequent failures of the pumps. High frequency random vibration was shown to be a good indicator for flow induced vibrations and cavitation detection. Cavitation and flow excited vibration on the pump with frequent unscheduled failures (pump SVP1) was successfully detected with vibration-based diagnosis. The failure modes in the pump showed shearing of the impeller blade at its root and cavitation erosion. Excessive clearance and/or journal bearing wear in the bad pump SVP1 also suggested consequential damage in this pump arising from higher vibration severity of the pump. The measurements also confirmed a better condition of SVP2 as compared with the pump of concern SVP1 consistent with visual examination of the physical conditions of the respective impeller conditions. This work also confirms that an assessment of the pumps operating conditions and in particular onset and/or occurrence of flow turbulence and cavitation could be made using vibration analysis in situations with inadequate pressure and flow instrumentations installed in pumps often being the case in typical industrial installations.

8 REFERENCES

[1] T.Uchiyama, S.Kallweit, & H.Siekmann (1998), "Fault Diagnostic Method for Pump Running Condition Based on Process Modeling and Neural Network", *International Journal of Rotating Machinery*, Vol.4 No.1 pp 49-59.

[2] M.Cudina (2003), "Detection of Cavitation Phenomenon in a Centrifugal Pump Using Audible Sound", Mechanical Systems and Signal Processing, 17(6) 1335-1347.

[3] Jorge L Parrondo, Sandra Velarde & Carlos Santolaria, "Development of a Predictive Maintenance System for a Centrifugal Pump", *Journal of Quality in Maintenance Engineering*, Vol.4, No.3 pp 198-211.

[4] Jeremy Jensen & Kenwood Dayton (2000), "Detecting Cavitation in Centrifugal Pumps", ORBIT, Vol. 21 No. 2, 2000, pp 26-30.

[5] Tan Chek Zin, M.Salman Leong (2005), " An Experimental Study on the Detection of Cavitation in a Centrifugal Pump Using FFT Vibration Analysis", 11th Asia Pacific Vibration Conference, Langkawi, Malaysia, November 2005.

[6] Maurice A.Y.(1996), " Pump Performance Monitoring", 5th International Conference on Profitable Condition Monitoring.

[7] Sulzer Brother Ltd. (1989), "Sulzer Centrifugal Pump Handbook", New York: Elsevier Science Publishing Co. Inc.

[8] B.K.Gandhi, D.P.S Chauhan, and M.M Panday (2002), "Diagnosing Cavitation in Centrifugal Pump using Noise and Vibration Signatures", 2th World Engineering Congress, July 2002, Malaysia : Sarawak. 325-330.

[9] Dieter-Heinz Hellmann (2002), "Early Fault Detection – An Overview", 9th International Symposium on Transport Phenomena and Dynamics of Rotating Machinery, Honolulu, Hawaii, February 2002.

9 ACKNOWLEDGEMENTS

The authors hereby acknowledge the assistance of the Independent Power Producer (IPP) where the vibration analysis was undertaken. This initiative allowed field measurements to be undertaken being the subject matter of this paper for validation of an experimental study in Universiti Teknologi Malaysia.

ASSET MANAGEMENT OF POWER APPARATUS

S. Birlasekaran and G. Ledwich

School of Engineering Systems, Queensland University of Technology, Brisbane 4001, Australia.

Abstract: This paper presents the details and difficulties on asset management process of power apparatus used in power network. The actual condition of power apparatus is far more important than its age. The regulatory and market forces in the current competitive energy market that are challenging the electric utility require more proactive methods of utility asset management. The utilities try to implement from the classical time based maintenance to a condition-based, predictive maintenance program using the concepts of reliability centered maintenance. Many diagnostic and condition monitoring techniques are developed for in-service power apparatus. Precise information about particular aspects of an assets performance is presented in a user-friendly way. Management group monitors the asset and network performance, investment and operating costs, and develop policy and standards, leading to an overall strategy for the network. On line static and dynamic data is the essential ingredient to effective asset management. The cost of condition monitoring can be recovered by reducing maintenance costs. Life cycle cost of an apparatus can be broken down into three distinct areas: purchase cost with installation, operations and maintenance, and decommissioning. The essence of asset management is to reduce the cost of keeping the asset in service and extending the period for which the asset provides satisfactory service. Assets age differ at different rates depending on the duty imposed on them. The current technology suggests that effective reliable age can be predicted more scientifically by blending financial aspects and condition indicators. The trend of condition indicators can predict potential failures and enabling corrective maintenance which is normally less expensive than repair following a catastrophic failure. More organized knowledge base needs to be developed to achieve this goal. This paper reviews the asset management process, the simple theory of power apparatus failure and the existing condition monitoring methods to predict that and the modern trend of intelligent asset decision process.

Keywords: Asset management, diagnostic and condition monitoring techniques, predictive and reliability centered maintenance,

1 INTRODUCTION

Management of aged and new optimally designed operating apparatus is the challenge to be faced by the utility managers to get reliable and economical electrical output in the competitive energy market. An asset manager has to make a variety of technical and managerial activities such as strategy and policy definition, fault management, asset value history and location, planning tools (system capacity, security, losses), risk management (system security, critical asset), system performance, health & safety policies, customer information and financial performance. Intelligent managers' move from routine time based maintenance to condition based maintenance to optimize the cost of operation. Condition indicators are being used to fix the remaining life of operating apparatus and to estimate the loading level.

The classical international standard [1] uses the thermal rule to identify the degradation level - like 10° rise in operating temperature may reduce life by half based on the laboratory measurements. The indicator of life may be tensile strength or breakdown strength which will reduce with increase in temperature of operation. In practical situation, apparatus may be subjected to linear and non-linear electrical, mechanical, chemical and thermal over loading at various levels. In addition, it operates in a power network of other aged apparatus in which the weak link has to be identified to cause the power outage [2]. To increase the reliability of power supply, operators move for on-line and off-line condition monitoring methods [3] [4]. Many of these methods are practiced depending on the expertise available in that organization. Mostly decisions are made from comparative measurements on similar type of aged apparatus. No precise universal rule can be formulated as a variety of conditions monitoring techniques are applied depending on the budget available for maintenance decision of 'go' or 'do not go'. If on-line monitoring is applied, the volume of collected data is more to store and process to come up with condensed useful information. Some techniques identify the degradation with the signal buried in the dominant background electrical and other noises, and ambient climatic disturbances [5] [6]. To get the dynamic condition indicators, signal processing and model studies are more effectively used with the improved low cost electronics and signal processing [7] [8]. This knowledge based data processing is the approach for modern asset management to identify the degradation rate, fault diagnosis and risk level [9] [10]. This paper presents some of our reported research work and some industrial practice for asset management of power apparatus.

2 ASSET MANAGEMENT

Asset management is a balancing act in which asset performance is balanced against shareholders expectations [11] [12]. Key stakeholders expect power companies to maximize earnings from existing assets while positioning themselves for future growth [10]. Reliability is defined as the probability that a system will perform a given function satisfactorily for a specified time under specified operating conditions. Reliability centered maintenance assures the delivery of safe, reliable electric power to customers. With that objective, asset management employs predictive modeling, risk management and optimized renewal decision making techniques to establish asset lifecycle options and related long-term cash-flow predictions. Tools exist to enable us to determine the condition of the plant. The asset management process adds value by converting this right information at the right time into decisions which reduce the overall lifecycle cost of the network. Fig.1 shows the flow of decision based on three requirements.

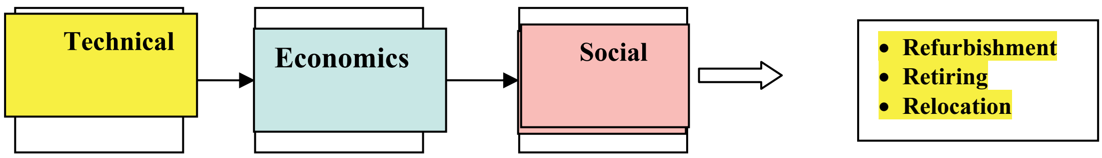


Figure 1 Management decision process

The technical aspect of the apparatus should be evaluated based on the various condition indicators, construction properties of that apparatus and the loading requirements.

The second factor is economic, based on assessing the total cost. It includes the present value of the apparatus (depreciated), the cost of replacement, cost of energy not supplied or cost of supplying the required energy using another source, corrective maintenance cost, preventive maintenance cost, operating costs of cost of energy losses, cost associated with technical condition.

The third factor is social and relies on risk management and optimal strategy with respect to essential requirement for human welfare. It evaluates the impact of any failure and acceptability of power interruptions in such locations. It depends on the critical locations like hospital and airport.

The last two factors can be evaluated by the application and financial managers. The technical factor is the important parameter in asset management to increase the asset value significantly. Many times, the asset can be refurbished by removing the causes of local initiating failure mechanism at a very earlier stage. To do that, the knowledge base is to be improved relating the trend of degradation level with the condition indicators. The condition indicators may be from mechanical, chemical, physical and electrical sensors and the range of fault frequency spectrum may vary. In this paper, some of the electrical condition monitoring methods are presented drawing upon case studies from the published literature.

3 THEORY OF FAILURE

In service, assessment of the age of an asset is not simple. Degradation of insulation is initiated in a localized area due to high stress. It spreads, leading to failure of the apparatus insulation and other supporting structures of windings.

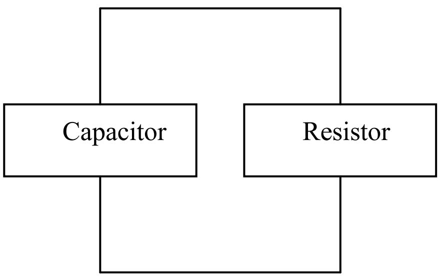


Figure 2 Model of power apparatus

This degradation is a slow process and the time from initiation of defects to failure may vary from one month to a few years depending on the stress. An ideal insulator is represented as a capacitor. Once the defects of different form built up, energy dissipation occurs and can be represented as a resistive element as shown in Figure 2. This is the electrical model of any power apparatus insulation in macroscopic view. The released active energy will be small and difficult to detect, a number of different condition monitoring diagnostic methods are available for use. Some managers use the absolute or relative value of condition indicator for taking some decision and many others rely on the rate of change of that condition indicator with time to make a decision. Normally, the confirmation of defect formation is done in a few stages. A good field engineer uses his common senses (like smell, visual inspection of change in color, sound and heat emitted by apparatus) related to load current with the installed preventive indicators like hot-spot indicator and Buchholtz relay operation [13]. The unusual operation will be taken as prelude for the initiation of defect in the system. Figure 3 shows the typical distribution of classified faults in transformers.

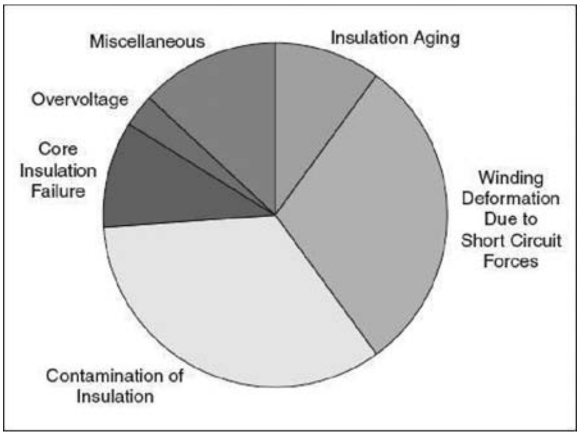


Figure 3 Distribution of faults in power transformers of 15 to 20 years old [14]

4 CONDITION MONITORING TECHNIQUES

Once the engineer identifies some unusual operation, one normally plans the preventive maintenance steps to avoid catastrophic failure. The corrective process once degradation is suspected is done in many ways depending on the expertise and tools available. It is traditionally done in off-line mode in which the power apparatus will be deenergised. Using external perturbation voltages, the developing defect severity and location can be identified. In modern days, on-line intrusive and non-intrusive techniques are effectively being used.

4.1 On-line diagnostic techniques

In this technique, the emitted signals due to normal 50 Hz high voltage (HV) operation are monitored to identify the faults.

4.1.1 Non-intrusive methods

4.1.1.1 **Thermography** – Heat generates infra red radiation. Using infrared camera at a suitable safe distance, the heated location of power apparatus can be identified [15] [16]. Figure 4 shows the change in color intensity at the heated location of joints in 3 Φ connection of a phase in transformer and circuit breaker. It is a non-invasive technique.

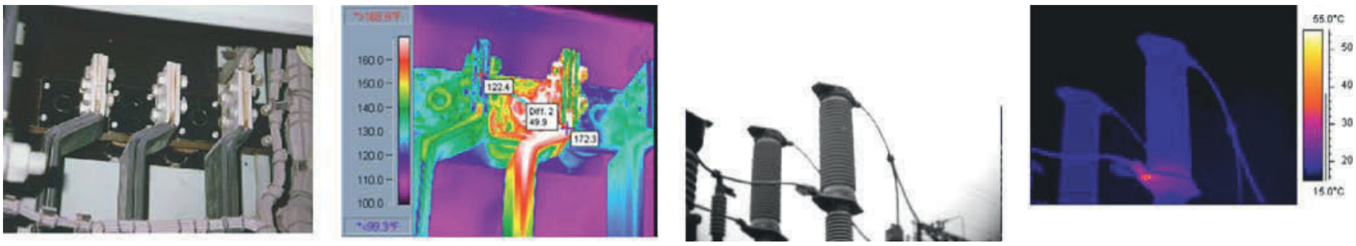


Figure 4 Thermography of faults in one phase of 3Φ connection [16]

4.1.1.2 **Partial Discharge Measurement** – External corona discharge or some form of exposed surface discharge due to contamination can be monitored by using ultrasonic or high frequency current transformer (HFCT) or transient earth voltage probe or antenna sensors in non-intrusive way. A typical low cost HFCTs and the measured PD signals in 20 ms period with 5 sampled data is shown below in Figure 5.

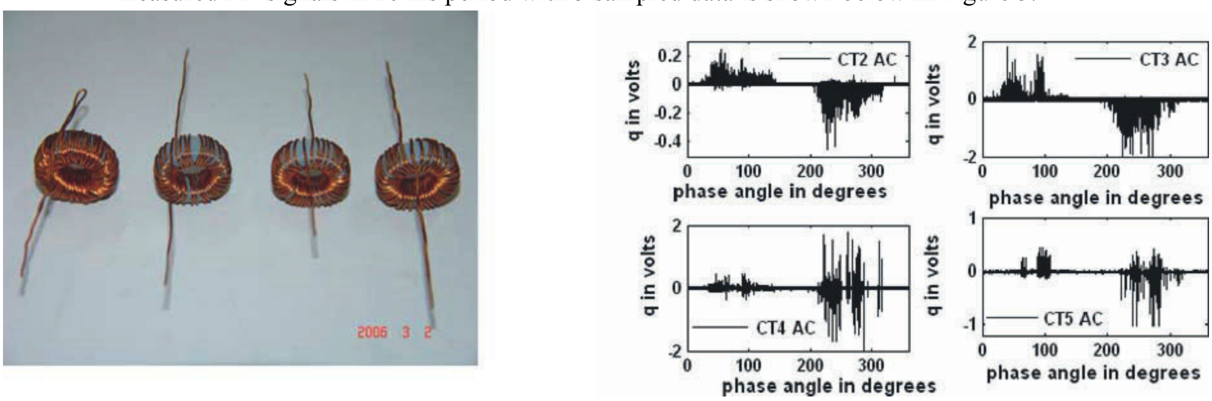


Figure 5 Fabricated 20 MHz HFCTs and its response to air corona [17]

4.1.1.3 **Loss angle measurement** – The leakage current through insulation of bushing is measured and then the ratio of in phase current (Equivalent current thro' resistor in Figure 2) and the quarture capacitive current (Equivalent current thro' capacitor in Figure 2) is determined continuously at 50 Hz operation. If the value exceeds 0.001, the unit will be tested with other additional techniques.

4.1.1.4 **Current measurement** – Load current is continuously monitored to relate with the heating. In addition, it is used to determine zero sequence impedance to identify the earth continuity of the circuit like corroded amour bonds and damaged leads.

4.1.1.5 **Vibration analysis** – Piezo electric sensors mounted at 3 or more locations of power apparatus can identify any non-symmetrical directional operation especially in rotating machine and will help to locate the faults in operating unit in on-line mode. They are widely used in motors and hydraulically /pneumatic operated circuit breakers.

4.1.2 Intrusive methods

4.1.2.1 Electrical methods

In intrusive mode, the measuring sensor will be placed near the HV terminal. It will transmit only the fault high frequency discharges but isolates the very high voltage 50 Hz component. Typical examples are the HV coupling capacitors installed near the line end terminals of generator of 11 to 22 kV rating or slot coupler which transmits the change in resistance (temperature in the slot) in the form of radio-frequency signals. Temperature sensors are mounted near the transformer top end of core and anticipated hot-spot locations. Distributed temperature sensing is done using fibre optics technique using the refraction changes and time domain reflectometry techniques. Pressure changes in long oil/gas -filled cable system are also monitored to identify the faulty location.

Since the signals are obtained without much attenuation, the sensitivity of detection will be high. But the cost of installing the HV capacitor will be considerable.

4.1.2.2 Chemical methods

In oil (or SF_6 gas) insulated system, samples will be taken in from operating transformer and it will be analysed for dissolved gas analysis, moisture content, furan content to relate to degree of polymerisation (DP) of paper, oil dielectric strength, and colour [13] [18]. It integrates the developing faults over time in the form of chemical changes. From the analysed pattern, the type of faults can be identified.

4.2 Off -line diagnostic techniques

Off line condition monitoring techniques have evolved over a period of more than 50 years. It can be classified into destructive and non-destructive measurements.

4.2.1 Destructive method

In this technique, test voltages will be destructive HV in the range of operating voltage.

4.2.1.1 **Hi-pot test -** In the initial period, engineers were using it as an acceptance test formulated in the International Standards [19]. An AC voltage in the range of 1.2 to 1.5 times the rated phase to phase voltage known as HiPot test - is applied to the isolated apparatus using external controllable 50 Hz HV source. The apparatus has to withstand one minute withstand voltage without any flashover or tripping. Since it is destructive test, step-stress Hi-pot test is applied for the aged apparatus. The test voltage is raised at 1 kV step and kept for one minute duration at that level before rising to next 1 kV up-level. Any abrupt rise in current with increase in voltage will indicate the degraded insulation and the increase in applied voltage will be stopped.

4.2.1.2 **Power factor tip-up test** – This is done by applying 50 Hz voltage in steps like step-stress Hi-pot test. Change in power factor or loss angle is measured with increase in voltage. When the ionization starts in the insulation, linearity will change and increase in loss factor will be noticed. Conventional Schering bridge is used to measure the variation.

4.2.1.3 **Partial Discharge test -** Variation of emitted partial discharge (PD) peak magnitude and number with increase in voltage is going to indicate the range of inception voltage to degrade the insulation.

4.2.2 Non- destructive methods

This is the very safe method of identification of fault. It can be classified into DC techniques, AC techniques and Impulse techniques based on the perturbation voltage.

4.2.2.1 DC technique –

- **IR test -** It is basically a megger test. The insulation resistance (IR) of the apparatus is measured with the megger of appropriate voltage rating. It is expected to be more than 100 MΩ. The tested apparatus should be discharged completely as it is DC test. DC test will store interfacial charges with long time constants and it will destroy the insulation gradually if it is not discharged. Also, AC apparatus composite insulating materials will be stressed in different manner with DC voltage.

- **PI test -** An extension of this DC technique is known as the polarization index (PI) test [19] [20]. It is found that the actual model is not as shown in Figure 2. The relaxation/absorption current of defect is shown as one series RC element and the measured current will vary as shown in Figure 6. To identify this relaxation process, the ratio of leakage current at 10 minutes and 1minute is determined to determine the insulation resistance (IR) ratio. PI is the ratio of IR at 10 minutes to 1 minute. If it is greater than 2, the insulation will work satisfactorily without defects.

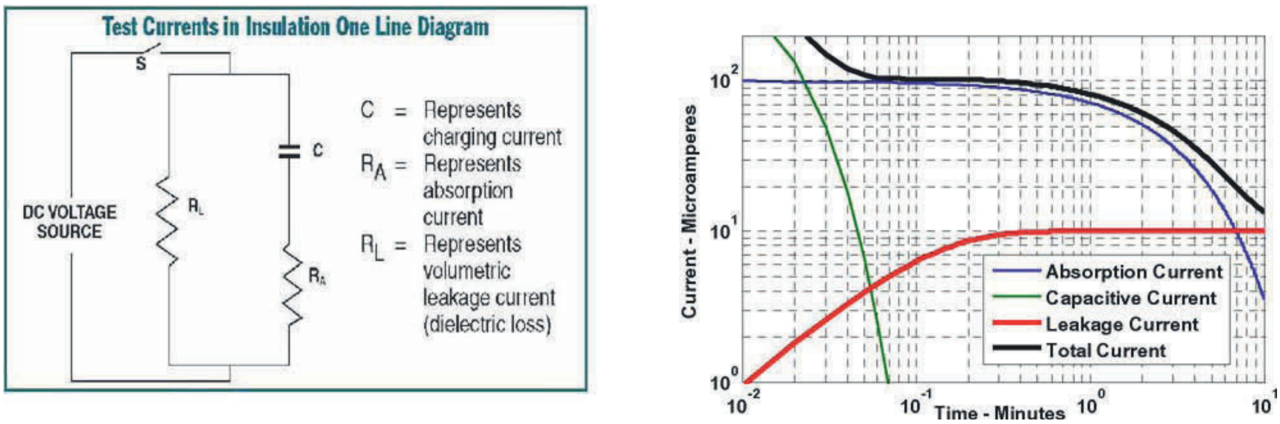


Figure 6 Improved Debye Model of apparatus [20]

Depending on the defects formed, many relaxation elements in the form of serially connected Rpi and Cpi will be existing in the insulation system [2] and the improved model is shown in Figure 7.

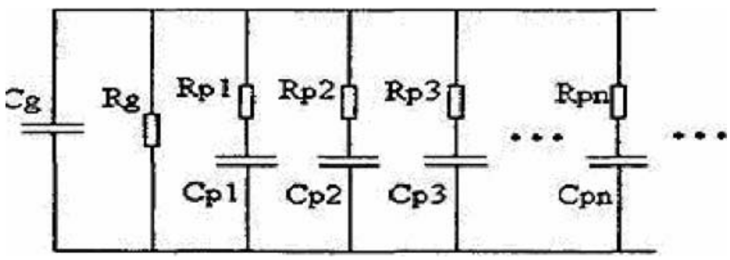


Figure 7 Improved Debye Model of apparatus [2]

- **PC/DP/IRC Test** – This test is known as polarization current (PC) or depolarization current (DPC) or isothermal relaxation current test [6]. Basically, a low DC voltage in the range of 50 V to 1000 V is applied to the system and the resulting polarization current with time will be measured. After that, the polarization source will be short-circuited ad the resulting depolarization current at uniform temperature will be measured. This depolarized current is released by the many relaxation elements shown in Figure 7. Some of them may have relaxation time up to many months. The measured current will be very low and careful measurement technique has to be applied under guarded electrode configuration. It is found to be very sensitive to detect the moisture infiltration into the insulation system.

- **AC technique** – By applying DC perturbation voltage, composite insulation of apparatus will be stressed differently. In addition, the polarized charge due to uni-polar voltage can stay in the interface of different insulations and will enhance the electrical stress at that localized site. To avoid the storage of charge in high insulation system with DC, AC technique is recommended as it simulates the real stressing condition and in each cycle, polarization and depolarization phenomenon occurs. This technique is an extension of Schering bridge technique. The difference is that the applied voltage will be low in the range of 50 to 200 V (rms) and the frequency will be changed.

- **Dielectrometry test** – In this method, the frequency is changed from 1 μHz to 1 kHz [21]. The resulting dielectric real and imaginary parameters are measured. As shown in Figure 7, both capacitance and resistance will change with frequency. At low frequency, a very high loss factor will be obtained to indicate the change in degradation significantly. It is very sensitive test and it requires very costly instrumentation system.
- **FRA test** – In this method, the structural changes in winding can be identified. The applied voltage is very low in the range of 1 to 3 V and the range of frequency used is from 1 kHz to 10 MHz [3][4][22]. A typical change in output is shown below in Figure 8. Significant signal processing and data base can be developed to identify the faults automatically. This technique may be developed to on-line method [3].

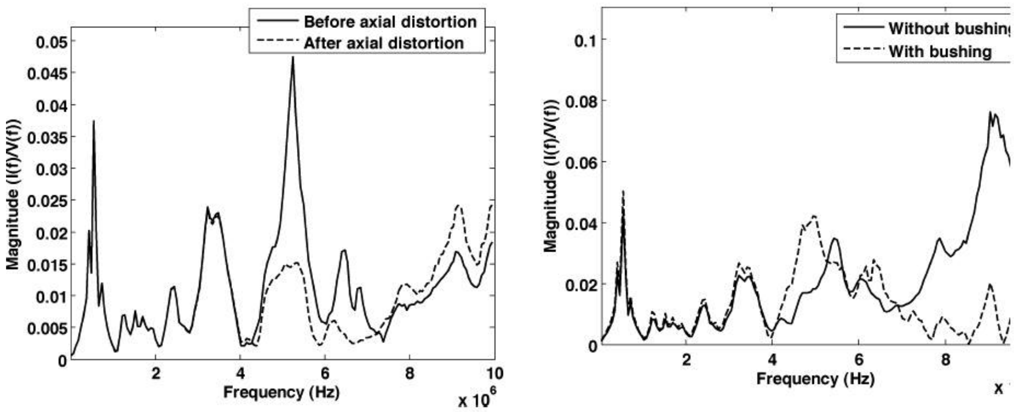


Figure 8 Frequency responses due to axial movement and effect of bushing [22]

4.2.2.2 Impulse technique

The impulse technique is applied in many ways.

- **PD test -** A discharge free 50 Hz HV source is applied to the insulation system. As the HV is increased, the partial discharge signal can be recorded as shown in Figure 4. This is random PD responses with other noises. The distribution of these pulses can be identified by cluster distribution in time and frequency plane as shown in Figure 9. The cluster group lying in the time range 0.6 to 0.8 μs and the frequency range 1.8 to 7.8 MHz is likely due to surface PD and other pulses are noises.

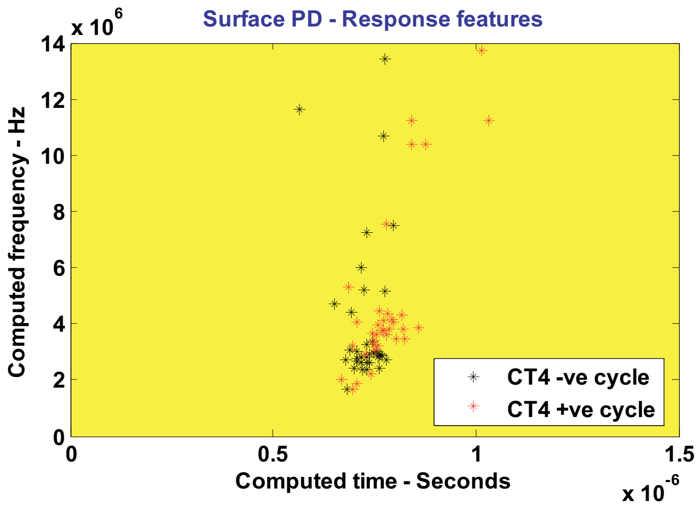
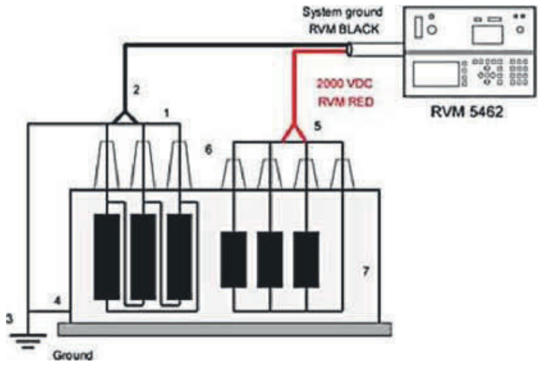


Figure 9 Random occurring single surface PD time and frequency distribution [17]

- **RVM Test -** In the industrial apparatus connected to power network, it is difficult to measure depolarization current due to many leakage paths. The new industrial technique known as recovery voltage method (RVM) is introduced [23] [24]. The apparatus will be charged to a finite interval like a step pulse of finite duration and will be discharged to less than the defined finite interval. After that, keeping the circuit open-circuited, the left-over charges stored in relaxation elements will be collected in the geometrical capacitor Cg in Figure 7. By varying the charging and the corresponding discharge interval, it is found that the recovered voltage (y-axis) reaches a maximum voltage at a certain time constant (x-axis) as shown in Figure 10. That time constant is found to be related to the moisture content and the remaining life is determined based on the time constant at which peak RV occurs. It is known as time domain technique to measure the relaxation behaviour.



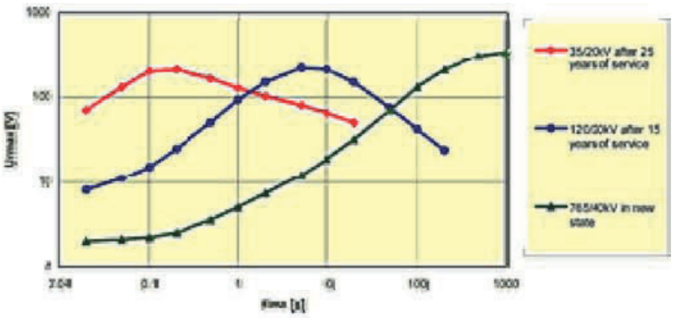


Figure 10 Recovery voltage measurement set up and typical responses [23]

- **LVI Test** – It is also a time domain test equivalent to FRA – frequency domain technique. A low voltage impulse in the range of 10 V with a double exponential pulse of fast rise time and slowly decaying time will be applied and the response current will be measured. Both signals will be transformed to get the transfer function response to identify the faults.

5 DISCUSSION AND CONCLUSION

Asset management of power industry is a billion dollar industry. Electricity has become a part of life these days in all parts of the world and one may not be able to visualize life without electricity, water and air. Recent blackouts in USA, Europe and energy demand in China [25] are a good example of need of more advanced asset management. Even in that assessment management process, social factors will play a dominant role. In a growing urban environment, medical, water, residential housing and transportation industries are looking for alternate energy sources to improve the reliability of electrical power supply. In the competitive energy market, economic factor is the dominant factor. In that price fixation model, models incorporating reliability factors associated with aged apparatus, the effect of failures in power quality and continuity of power supply in network have to be developed.

In the modern information era with good communication facility, static and dynamic technical conditions of operating apparatus can be obtained using web network without visiting any site. Many of the starting failure processes occur in a localized high stress volume of the apparatus. Normally, the failure time from inception of degradation to failure of operation may vary from one month to a few years. Section 4 clearly suggests that many of the condition monitoring techniques can be implemented easily at the site and the cost can be recovered in the form of increased asset life. With the reported condition monitoring methods and other mechanical, physical and chemical diagnosing methods, the developing fault can be identified. Most of the localised faults can be repaired. After repairing and knowing the conditional status of the apparatus, the unit can be put into use again or moved to another location with less demanding load. By this way, the effective use of asset can be extended by knowing the conditional status [26]. The life (book value) of many of the apparatus varies from 15 to 25 years in economical value. But many utilities continue to use the aged apparatus by knowing the operational condition of such apparatus. The knowledge of identifying the conditions and the recorded variation of different critical parameters will become a knowledge base for taking any decision. At present, no steps are taken in the decision process to incorporate it as a static knowledge base since they are not collected and stored properly.

To make a scientific assessment process, knowledge base of different condition indicators indicated in section 4 and by other methods has to be built up. Dynamic instrumentation with good signal processing facilities has to be developed. Dynamic model of the aging apparatus is to be developed for reliable prediction of failure.

Many experienced engineers are retiring with the observed knowledge base of specialist skills and a smart system needs to be developed to collect such human- resource knowledge to blend with ever developing technology, brain-power and new high quality materials used in different modern apparatus [9].

6 REFERENCES

[1] IEEE Std. 98 (2002) Standard for the preparation of test procedures for the thermal evaluation of solid electrical insulating materials.

[2] Birlasekaran S & Xingzhou Y. (2003) Relaxation studies on power equipment, IEEE Trans. On DEI, Vol.10 (6), 1061- 1076.

[3] Birlasekaran S & Fetherston F. (1999) Off/on-line FRA condition monitoring technique for power transformer, IEEE Power Engineering Review, 54-56.

[4] Birlasekaran S, Xingzhou Y, Fetherstone F, Abell R & Middleton R. (2000) Diagnosis and Identification of Transformer Faults from frequency response data, Proc. Of IEEE Power Engineering Society Winter Meeting, Vol.3, 2251- 2256.

[5] Birlasekaran S. (2003) Identification of the type of partial discharges in an operating 16 kV/ 250 MVA generator, Proc. Of Conf. on Elec. Insulation and Dielectric phenomena, 559-562.

[6] Heifang S, Birlasekaran S & Xiaolong C. (2000) Diagnosis of power cable insulation using isothermal relaxation current technique, Proc. Of the 6th Int. Conf. on Properties and Applications of Dielectric Materials, 411- 414.

[7] Ming Y, Birlasekaran S & Choi SS. (2005) Simulation of partial discharge propagation in power network, IEEE Trans. On energy conversion, Vol.20 (3), 644- 652.

[8] Birlasekaran S & Ledwich G. (1998) **Use of FFT and ANN techniques in monitoring of transformer fault gases**, Proc. Of Int. Symp. On Electrical Insulating Materials, 75-78.

[9] Allan DM. (2005) Asset management challenges in aging power systems, Proc. Of Int. Symp. On Elec. Insul. Materials, Vol.1, 265-268.

[10] Shahidehpour M & Ferrero R. (2005) Time management for assets – Chronological Strategies for Power System Asset Management, IEE Power & energy magazine, 33-38.

[11] Quak B, Gulski E, Wester Ph & Smit JJ. (2003) Fundamental aspects of information processing and the decision process to support asset management, Proc. Of 7th Int. Conf. on properties and applications of dielectric materials, Vol.3, 982-985.

[12] Muhr M. (2005) Asset and risk management of electrical power equipment, Proc. Of Int. Symp. On Electrical insulating materials, Vol.3, 861-864.

[13] Emsley AM & Stevens GC. (1994) Review of chemical indicators of degradation of cellulosic electrical paper insulation in oil-filled transformers, IEE Proc. On Science, Measurement and Technology, Vol.141 (5), 324-334.

[14] Wang M, Vandermaar AJ & Srivastava KD. (2002) Review of condition assessment of power transformers in service, IEEE Trans. EI Magazine, Vol.18 (6), 12-25.

[15] Niancang (1998) The infrared thermography diagnostic technique of high-voltage electrical equipments with internal faults, Proc. Of Int. Conf. on Power System Technology, Vol.1, 110-115.

[16] www.flirthermography.com.

[17] Birlasekaran S & Dongcheng C. (2006) Comparison of known PD signals with the developed and commercial HFCT sensors, Paper sent to IEEE Trans. On Power Delivery.

[18] Allan DM. (1993) Practical life-assessment technique for aged transformer insulation, IEE Proc. On Science, Measurement and Technology, Vol. 140(5), 404-408.

[19] Stone GC (2005) Recent important changes in IEEE motor and generator winding insulation diagnostic testing standards, IEEE Trans. On Industry Applications, Vol.41 (1), 91-100.

[20] www.aemc.com/techinfo/techworkbooks/ megohmmeters/tech_megohm.pdf

[21] Du Y, Zahn M, Altamirano N, Sarda M, Mamishev AV & Lesieutre BC (2002) Moisture and temperature effects on the dielectric spectrum of transformer pressboard, Proc. Of Elec. Insulation and Dielectric Phenomena, 878-881.

[22] Wang M, Vandermaar AJ and Srivastava KD. (2005) Improved detection of power transformer winding movement by extending the FRA high frequency range, IEEE Trans. On Power Delivery, Vol.20 (3), 1930-1938.

[23] www.high-voltage-hubbell.com/ products/5462/5462body1.htm.

[24] Cherukupalli S, Buchholz V, Colwell M, Crine JP & Keefe RJ (2004) Condition assessment of distribution PILC cables from electrical, chemical, and dielectric measurements, IEEE Elec. Insulation Magazine, Vol.20(4), 6-12.

[25] Lin-Liu J & Sweet W (2004) Energy tangle in china, IEEE Spectrum, Vol. 41(10), 12-14.

[26] Arshad M, Islam AM & Khaliq A. (2004) Power transformer asset management, Proc. Of PowerCon – Power System Technology, Vol.2, 1395-1398.

REMOTE SENSING IN CONDITION MONITORING

R Pietruszkiewicz[1], A G Starr[1], A Albarbar[1], R Mattheys[2]

[1] *Through Life Support Research Group, School of Mechanical, Aerospace and Civil Engineering, The University of Manchester, Sackville Street, Manchester M60 1QD, UK*

[2] *Engineering Statistical Services Ltd, 51 Twizell Lane, West Pelton, Stanley, County Durham DH9 6SQ, UK*

Abstract: This paper reviews the latest trends in the use of intelligent sensors in monitoring applications, including new approaches and computation methods. Sources used emphasise the scientific motivation bringing technological development in the monitoring area, as well as commercial and technological trends.

Key points reviewed include:

- **Parameters and types of sensors currently used**
- **Scientific interest in intelligent sensors**
- **Currently developments in applications in academia**
- **Benefits from using sensors**
- **Processing capacities offered by intelligent sensors**

The state of the art in intelligent sensors capitalises on novel technologies which perform monitoring functions robustly, inexpensively, and at extremely low power.

Key Words: remote sensing, distributed intelligence, wireless, power harvesting

1 INTRODUCTION

The categorisation of a sensor as “intelligent” implies that the sensor incorporates more functionality than merely providing an output measurement. There is some discrepancy governing what makes a given sensor intelligent. There are varying levels of sophistication used by sensors, which have claimed to be intelligent or smart, ranging from merely incorporating an operational amplifier for the output signals, to advanced data modelling techniques for condition monitoring. A “smart” sensor is defined as a sensor that provides functions beyond those necessary for generating a correct representation of a sensed or controlled quantity and having on board advanced processing capability. The intelligent sensors discussed here have built in wireless communication capability. Features required for improving confidence in sensor measurements are defined. This function typically simplifies the integration of the transducer into applications in a networked environment.

The paper introduces possible measurements and available sensors, discusses intelligent sensors in research applications, explaining benefits and advanced possibilities, and finally reviews communication standards used in wireless technologies.

2 PARAMETERS AND SENSORS IN CURRENT USE

In research for the latest applications of intelligent sensors, it will be beneficial to start with an outlook on the applications of conventional sensors. Table 1 lists fields of interest from a scientific point of view [1].

Their applications include:

- "electronic nose" technology, voting systems, telemetry;
- automotive, medical, environmental monitoring and control, consumer, alarm and security, military, nautical, aeronautical and space, robotics and automation;
- sensitive skin systems, sensor-actuators.

Some of the above mentioned applications were designed as intelligent, remotely accessible devices. Development of the technologies introduces a broader need for incorporation of intelligence and remote access. It is expected that in the near future most conventional sensors will be available with upgraded intelligence built into their structure. Such a framework is generic, with the intention of being able to apply the upgraded intelligence to a variety of sensor types, and across civil and military sensor platforms. Standardisation is important to improve portability, and to reduce cost.

Strategies for improving functionality are important, as are features of the sensors themselves, including:

- integration, such as multiple-sensor systems, intelligent sensor arrays (large and high density), sensor buses and distributed networks;
- combined sensors (e.g., electrical & mechanical);
- sensor phenomena and characterization (sensitivity, selectivity, noise, aging, hysteresis, dynamic range, and interference effects).

Table 1: Sensing parameters in intelligent sensing [adapted from 1]

- Mechanical devices and parameters such as: metallic, thin-film, thick film and bulk strain gauges, pressure, accelerometers, angular rate, displacement, force, bulk and surface acoustic waves, ultrasonics, flow;
- Electromechanical sensors over a wide measurement and substrate range;
- Thermal sensors such as: platinum resistors, thermistors, diode and transistors, thermocouples, thermopiles, pyroelectric and piezoelectric thermometers, calorimeters, bolometers;
- Optoelectronic/photonic sensors such as: photovoltaic diodes, photoconductors, photodiodes, phototransistors, position-sensitive photodetectors, photodiode arrays, charge-coupled devices, light-emitting diodes, diode lasers, other quantum devices, liquid-crystal displays;
- Ionizing radiation sensors such as gamma ray, charged particle and neutron detectors;
- Integrated optics/fibre optical devices such as those based on photometry, fluorimetry, interferometry and ellipsometry;
- Microwave/millimetre wave sensors;
- Magnetic sensors such as: magneto resistors, Hall-effect devices, magnetometers, magnetic-field sensors, solid-state read and write heads;
- Chemical and biological sensors, with emphasis on the electronics and physics aspects of transducing chemical and biological signals into information about chemical and biological agents;
- Mass-sensitive devices such as quartz crystal microbalances and surface acoustic wave devices;
- Other sensors such as: thin-film and thick-film gas sensors, humidity sensors, ion sensors (such as pH sensors), radon sensors, carbon monoxide sensors, viscosity sensors, density sensors, acoustic velocity sensors, proximity sensors, altimeters, and barometers.

3 SCIENTIFIC INTEREST IN INTELLIGENT SENSORS

The main interest for the use of wireless intelligent sensors comes from the advanced functionality offered by this type of sensor. Unlike conventional sensors, they are designed to create a complex architecture enabling implementation of advanced measurement strategies. This facility is not present, or is relatively difficult to establish, with conventional sensors. Intelligent sensors do not measure new parameters: the use of them allows measurement of the same parameters but on a different scale. Applications of wireless networks add a new level of sophistication in feature extraction, allowing better monitoring than was previously possible. Additionally, use of distributed intelligence, in the form of local data processing, allows for faster sampling, and reduction of the retained data.

Sensor networks represent a significant improvement over traditional sensors. They are deployed in the following ways:

- Sensors can be positioned far from the actual phenomenon, for example something known by sense perception. In this approach, large sensors that use some complex techniques to distinguish the targets from environmental noise are required. This is in case where the specific measurement is not the main point of concern, but the effects are. For example in the building automation, the wind force will have effect on the stability of the building.
- Several sensors that perform only sensing can be deployed. The positions of the sensors and communications topology are carefully engineered. They transmit time series of the sensed phenomenon to the central nodes where computations are performed and data are fused. This is especially in case of the complex machinery or large objects requiring multiple points of measurement.

An intelligent sensor approach can also be used to improve confidence in the reliability of sensory data in the higher-level sensor management. Onboard condition monitoring and fault detection techniques can be used in preference to reliance on sensor redundancy for ensuring robust measurements. Advanced data based modelling techniques can be used to model non-linear and time-variant sensor systems, avoiding the limitations of linear physical sensor models, and allowing for reconfiguration of the sensor to correct possible errors.

Other functions of an intelligent sensor are autonomous capability, adaptation to changes in the environment, and tolerance to faults.

The principal characteristic of such an intelligent sensor is that it is capable of communicating reliable and self-validated signals or features to higher-level supervision systems, for purposes such as information fusion, tracking and estimation. Poor sensory data can be identified by the sensor itself and flagged for quality problems, together with estimates of the likely cause to enable attempts at sensor reconfiguration. [17]

4 APPLICATIONS IN SCIENTIFIC AREAS

Many research areas benefit from the implementation of intelligent sensors. Applications where wireless sensors have been used range from vibration monitoring, through detection of micro-organisms in food products by sensing the temperature rise of the products, to capacitive displacement transducers which can accurately sense position, speed and acceleration by measuring the (trans)capacitances in multi-electrode structures. Neural networks for local processing are finding many uses, particularly in multi-element sensor arrays such as those found in the "electronic nose".

The range of applications for wireless sensors is very large. A few examples from the literature are presented in Table 2.

Table 2: Wireless sensing applications

- Structural health monitoring (SHM)
  - Vibration measurements are applied for condition monitoring of machines and structures, structural integrity assessment, damage detection and structural failure prediction. Multi-sensor measurement data is essential for SHM. Application of many sensors located on mechanical structures without cables makes the monitoring process more efficient. [2]
  - A chloride sensor and a radio-frequency identification (RFID) chip can be queried remotely both to identify it and to indicate chloride concentration levels; [3]
  - Integrated wireless piezoelectric sensors: Piezoelectric arrays and networks have been suggested as a means to monitor the integrity of composite structures throughout the service life, for instance of an aircraft; [4]
  - Noise reduction in RF cavity wireless strain sensors: monitoring strain in civil structures; [5]
- Passive wireless strain and temperature sensors: approaches to wireless strain and temperature measurements that employ passive sensors based on two types of surface acoustic wave devices, reflective delay lines and resonators; [6]
- Surface acoustic wave devices based wireless measurement platform for sensors: in some applications, a wireless readout is necessary because of the difficulty of fixed connection between sensors and signal processing unit. Wireless surface acoustic wave sensors are being used for sensing some physical and chemical phenomena passively; [7]
- Gaps measurement: a network of small, wireless sensors is helping measure seal gaps in real time; [8]
- Nano-scale gas sensors: microwave carbon nanotube; [9]
- Damage detection and correlation-based localization: demonstration using wireless sensors; [10]
- Passive wireless surface acoustic wave (SAW) sensors based on Fourier transform: application for different physical parameters such as temperature, pressure, and force; [11]
- Continuous monitoring of vital bio-signals; advanced monitor and control medium for healthcare services; [12]
- Agriculture and food production for environmental monitoring, precision agriculture, man-to-machine based machine and process control, building and facility automation and RFID-based traceability systems; [13]
- Microwave moisture sensors: battery powered and capable of communication with the host control system via spread spectrum wireless communications. [14]
- Model analysis: Approaches to wireless hardware and software are suggested that could parallel calculations and thus reduce calculation time and improve data quality by elimination of wires. [15]
- Smart microphone, suitable for outdoor acoustic surveillance on robotic vehicles, incorporating MEMS sensors for acoustic sensing, wind noise flow turbulence sensing, platform vibration sensing. [16]

5 BENEFITS FROM USING INTELLIGENT SENSORS

There are direct benefits from using intelligent sensors. They are able to operate in a task-specific manner with effective data collection techniques. They enable the development and application of more flexible sensor networks which efficiently utilize and coordinate the limited resources of each individual sensor. By focusing resources according to the state of the surrounding environment and on the immediate task, more efficient operation of the sensor is ensured. [18]

- Accuracy: an intelligent sensor will incorporate features which enable it to compensate for systematic errors, system drift and random errors produced due to system parameters or the characteristics of the sensor.
- Reliability: the incorporation of data and sensor validation techniques to detect corrupted data, self-testing of network path connections and sensor operation, as well as calibration of sensor drift, provides yet another level of system reliability in addition to techniques already applied in the network design.
- Adaptability: the processing parameters of an intelligent sensor system should be determined automatically and adopted by a higher level in the system architecture. This enables the optimization of the measuring and processing operations, as well as enabling the sensor to adequately respond to changing environmental conditions.
- Bandwidth reduction: in the case where the number of sensors in a system is expanding, it might cause severe data processing bottlenecks. By localising signal processing and reduction, the data communication bandwidth can also be reduced. [19]
- Advanced data processing: enabling the "intelligence" to be implemented to the sensing strategy - this is perhaps the main advantage of using intelligent sensors above conventional sensors.

6 PROCESSING CAPACITY AND ARCHITECTURE IN INTELLIGENT SENSORS

An important advantage of intelligent sensors is the distribution of advanced processing methods to the point of measurement [20]. The capability of microprocessors is important for performance in smart devices. Digital signal processors (DSPs) are running faster and have improved development tools, making it simpler for designers to employ them in applications which need faster mathematical processing than general-purpose microprocessors can provide. DSPs also make it possible to run more diagnostics. Distributed smart sensors can integrate algorithms which would normally be run on a supervisory system. Local memory allows reference to look-up tables and historical data, allowing adaptable strategies. Local processing allows rapid, frequent measurement establishing good characterisation of both healthy and faulty data without unnecessary network traffic. Fast processors, especially DSPs, are power hungry and this will be an important problem for wireless systems.

A second advantage of intelligent sensors is the opportunity of working together in small or large numbers, and only communicating the results if necessary. A sensor array may be used to enhance reliability by allowing for redundancy. This may detect faults in the sensors themselves and allocate tasks to adjacent sensors. Different analysis tasks may be distributed to different processors within the array, allowing a single system to perform several analyses simultaneously. Recognition and analysis algorithms might be distributed across the array as a whole, creating spatially-distributed sensory fields, which can be particularly advantageous for multivariate monitoring problems and data fusion. Pattern recognition has been widely used off-line for diagnosis of mechanical and system faults, and sensor networks allow an online approach. Systems are likely to use all these possibilities, providing systems with specialised processing and an element of redundancy, providing powerful diagnostic procedures with graceful degradation in the case of individual sensor failure [19,24].

Advanced processing methods, currently only employed by researchers, can be embedded in distributed systems, and this is a key strength of intelligent sensors above conventional devices. The methods available are briefly described in the following sections, and are applications which are described by the data fusion architecture, where the concern is to extract information from data, supported by knowledge, in order to make estimations, classifications or decisions [21].

6.1 Pre-processing: sub-object level

The physical sensing element or transducer produces a response to its environment. The pre-processing stage converts the transducer's response to its input (e.g. temperature) into a more useful engineering quantity which is representative of the raw environmental parameter, such as electric current or voltage. Pre-processing includes software for calibration, which may include linearization, normalisation, and conditioning of the signal to correct for deviations, e.g. those caused by temperature effects or ageing in the sensor and its electronics. [17]

6.2 Information processing: object refinement

Information processing uses filtering and data manipulation for estimation, feature extraction and classification to achieve data reduction and event detection, enhancing and interpreting the collected data [18]. Learning techniques such as fuzzy logic or neural networks may be used to produce a high confidence feature from weak data. Some applications need qualitative (symbolic) characterization of the estimate, for example for rule-based control or diagnostics, where the variables are described by qualitative values for example large, small, increasing, decreasing [23]. The aim is to present the right information to the next level of the hierarchy, with the minimum necessary data. This is an important architectural feature of distributed systems, allowing more local data to be acquired and interpreted, while reducing network traffic – most systems are healthy most of the time.

6.3 Condition monitoring and fault detection: situation and assessment

A condition monitoring system firstly detects change from normal, and then attempts to determine the location, type and severity of a fault. Ideally we could make an estimate of remaining life. Distributed sensors immediately aid the determination of fault location. Local processing, e.g. the fast Fourier transform for feature extraction of a vibration signal, can be used for an indicator of fault type [22]. It is possible to store a look-up table of previous or typical data, to which feature data can be compared to determine severity.

6.4 Sensor modelling, configuration and uncertainty: process refinement

The interpretation of a sensor output depends on the models described in 6.1 and 6.2 above. They are usually simple, linear relationships, and are usually assumed to be constant with time. The rich data acquired from fast, regular sampling allows a better approach to the refinement of models. Non-linear models, such as kernel representations, and other data-oriented models, can allow automatic retraining of the sensor, which may be used to deal with ageing effects. Uncertainty estimates can be returned to a higher level in the hierarchy, which is important for some data fusion algorithms [17]. It is also possible to download new models after comparison with other sensors' behaviour. Compensation may be made to account for variation [18].

6.5 Communications

Intelligent systems require contact with a wider system to report normality and faults, and to be programmed. Traditional sensors are often limited by a lack of standardisation, leading to complex systems, high cost and low flexibility. Standardized transducer interfaces and communications protocols, which are sophisticated in wireless systems, result in autonomous, distributed, re-configurable sensors which are linked in a coherent manner efficient communications and fault tolerance [18].

6.6 Fault Tolerance.

Monitoring of the sensor system itself requires a specially robust architecture to differentiate real parameter changes from sensor faults [17]. Errors arise from behaviour which does not match the expected models, and if they are not detected, they will be introduced into the monitoring system. Fault tolerance issues can be considered at different levels [23]. Real time strategies allow the system to continue its operation with faults present. The fault may be compensated so as to avoid erroneous estimates, or the sensor system may be reconfigured to provide the estimate in another way, e.g. using different sensors. It is also possible to evaluate the system fault tolerance i.e. its ability to accept faults while being still able to run. This depends on the nature of the fault: critical resources e.g. the power supply, can only be tolerated by hardware redundancy.

6.7 Validation

The introduction of erroneous data to the system could cause serious problems, on a wider scale than a single sensor failure. If the data is plausible but wrong, the wrong decision could be made for a component or system, such as allowing an aircraft to fly with an incipient fault. The incorporation of validation into intelligent sensors increases the reliability of the system. The impact of errors can be reduced by redundancy, or at least a dense network where nearby sensors can be compared. Hardware redundancy uses extra sensors, comparing data which appears similarly on the majority of sensors. Analytical redundancy uses a mathematical model to compare the relationship between sensor measurements, computing the difference between an estimate and actual, with either an observer or a residual approach. This can become computationally expensive as complexity increases. The validation can qualify or disqualify the estimation or at least which evaluate the confidence level which can be associated with it [18,23].

A partial data validation approach is technological validation, which checks hardware resources, including power supply checksum of the microprocessor’s memory, and connection to the network. This does not check the data estimation, just the operating conditions [23].

7 RADIO FREQUENCY (RF) STANDARDS

Narrow-band radio technologies are finally disappearing, having limitations for some implementations. Several new technologies are making rapid progress. IEEE 802.11 (WiFi) may have a big impact in industrial markets. Bluetooth technology is still an alternative and a common industrial communication standard. Ultra wide band (UWB), is new to the marketplace and could be a serious option in many applications. This has caused a number of vendors to abandon their own radio systems and cease internal radio development effort in favour of the huge cost and effort saving of low cost wireless standards. This adoption of standard solutions (often created with other markets in mind) has served condition monitoring well before, e.g. in portable computing and instrumentation.

Table 3 shows the existing communication technologies. A short comparison of the most popular technologies and their parameters is presented in Table 4, and their range is illustrated in Figure 1.

Bluetooth still needs consideration as a generic solution. The Bluetooth consortium recently approved an industrial automation working group to address the concerns of the marketplace. Some vendors find it inappropriate for their particular market and applications. There are tradeoffs between cost and reliability in Bluetooth, but the technology should not be discounted just because it is not ideal; it may still be successful. Two spin-offs originating from the Bluetooth standard are Bluetooth enhanced data rate (EDR), with a better bandwidth, and a new protocol in industrial communication, ZigBee. A number of companies, including several new on the market, are actively bringing wireless technologies to market.

8 POWER SOURCES

Power for sensors is important, but for intelligent wireless sensors it is critical. The applications and systems architecture in a distributed system are affected by the performance of the power supply. A typical application uses a battery, which brings a tight restriction on the power consumption and the life of the device without maintenance. To reduce the power consumption, the processor should have three sleep modes: idle, which just shuts the processor off; power save, which leaves an asynchronous timer running; and power down, which shuts everything off except the watch-dog. Power consumption is inversely proportional to battery life. In some applications, up to ten years life is required. Processing, radio, and a typical sensor load can easily consume over 100 mW – compare this with the 30 μW when all components are in sleep mode. The power usage strategy must expedite tasks before returning to sleep mode. This power usage strategy is a key feature of the infrastructure software [25].

Power harvesting opportunities are a natural trend to improve the lifespan and maintenance free time of the installed sensors. Some sensor applications require a high level of self-sufficiency from the device and the harvesting of ambient power enables them to operate without maintenance – specifically a battery changing operation. Alternatives for scavenging power include vibrations, acoustic or millimetre wave energy through use of sensor resonators or piezo-electric devices, solar power and electromagnetic fields. The options increase as the radio range and power consumption diminishes. Sources are compared in Table 5.

Table 3: Figure1: Wireless Standards for Data Application [37]

Standard		Mature	New	Trial	Development
ETSI	2G	GSM GPRS	EDGE		EDGE Ph2
	3GPP	UMTS/WCDMA	TD-CDMA HSDPA	HSUPA	
	3GPP2	CDMA	1xRTT	1xEV-DO	1XEV-DV
IEEE	802.11	Wi-Fi.11b .11a .11g			.11n
	802.15	802.15.3 – UWB		MBOA/DS-UWB	
	Bluetooth	802.15.1 v1.1	v1.2 v2.0+EDR	.15.4(ZigBee)	
	802.16	WiMAX	.16c	.16d	.16e
	802.20	MAN			.20

ETSI standards including: 2G, 3GPP, 3GPP2 are the technologies used in mobile phones.

IEEE standards containing the range 802, are low range wireless.

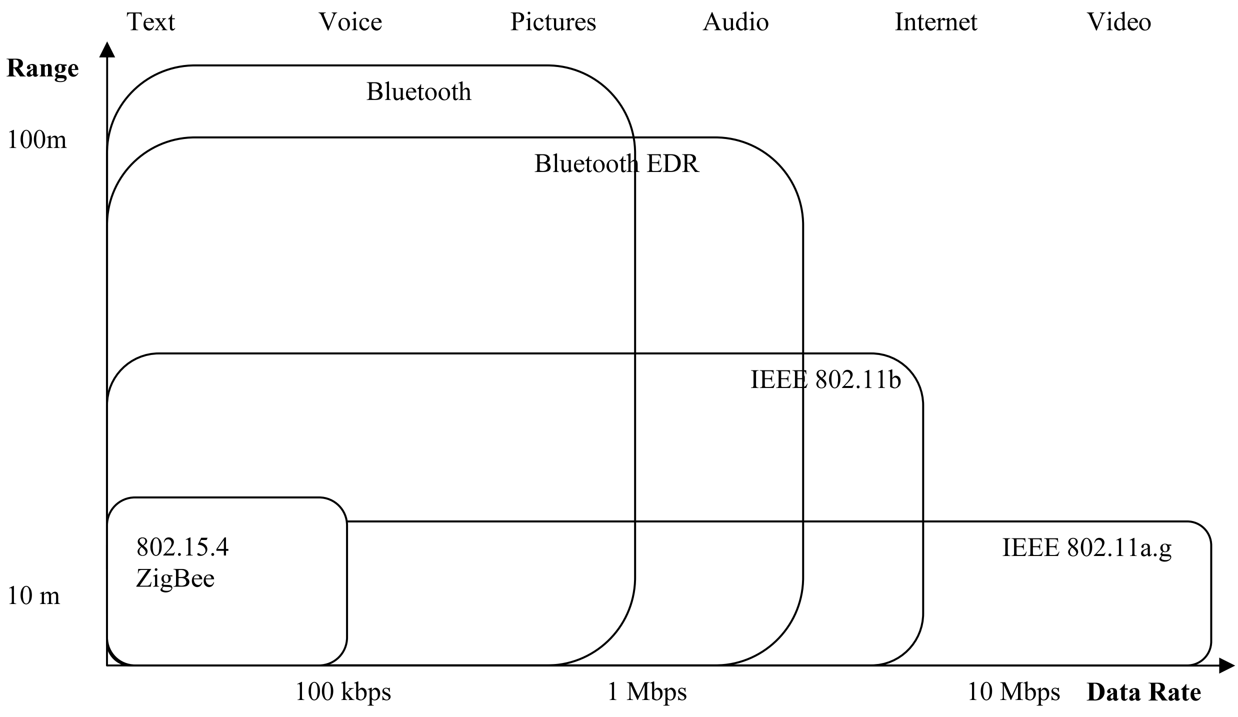


Figure 1: Range comparison at similar power levels [37]

Table 4: Comparison of Wireless Communications Protocols [36]

Parameter	Wi-Fi (IEEE 802.11)	Bluetooth (IEEE 802.15.1)	ZigBee (IEEE 802.15.4)
Range	About 50 m	About 10 to 100 m	About 10 m
Bandwidth/ Throughput	868 MHz/20 kbits/s 915 MHz/40 kbits/s 2.4 GHz/250 kbits/s	2.4 GHz/1 Mbit/s	868 MHz/20 kbits/s 915 MHz/40 kbits/s 2.4 GHz/250 kbits/s
Power consumption	400 mA (TX on) 20 mA (standby)	40 mA (TX on) 0.2 mA (standby)	30 mA (TX on) 1 uA (standby)
Protocol stack size	100 kbytes	100 kbytes	32 kbytes
Battery life	Minutes-hours	Hours-days	Days-years
Relative node physical size	Large	Medium	Small
Relative cost/ complexity	High	Moderate	Low

Table 5: Comparison of power Sources

Power source	Advantages	Disadvantages
Battery powered	• Long lasting source • Cheap solution • Reduced power management • Reliable technology	• Additional cost • Environmentally unfriendly • Limited/uncertain life span • Maintenance required
Online power supply	• No need for energy saving power management • Reduced cost of the device • Maintenance free • Reliable	• Cabling distance restrictions • Closeness to the source • No mobility possible • Reliable on the continuity of external supply
Power harvesting	• Maintenance free - No battery replacement • Enhance lifespan of device • Environmentally friendly • Modern approach with big prospects	• Complicated power management • Additional cost of device • Only applicable in right environment • Need for additional source or power storage when device was stopped

9 CONCLUSIONS

Over the past decade or so, intelligent sensors have risen from being an academic pipe dream into practical and commercial devices. The main reason for their continued success is largely as a result of the major advances in the area of microelectronic technologies. Limitations of conventional sensors can be compensated by the advantages brought by intelligent, miniaturised and wireless sensors.

Like any sensor, an intelligent sensor is primarily aimed at providing estimations of system variables, to be used in control loops, or in decision making for control, maintenance or management of the system. However, intelligent sensors are not components that are just interconnected to the rest of the application. An intelligent sensor should autonomously perform advanced processing functions such as self-validation, self re-calibration, fault detection, and sensory data filtering and feature extraction.

The presented functional view of intelligent sensors shows that features of interest from the scientific point of view are the growth in processing capabilities such as: estimation, characterization, validation and fault tolerance. On the contrary, due to the extra functions they implement, they fully participate in the architecture of complex distributed control systems by offering services at the supervision level.

One of the prime issues with wireless sensor networks is the power consumption of numerous sensor nodes and the requirements to provide periodic maintenance including battery replacement. This issue can be resolved by implementation of the energy harvesting methods, use of which can eliminate the need for battery replacement, making sensors even more independent and their functionality even more distributed.

Future research is likely to focus on advanced data processing directly in the sensor. True distributed data processing opens up new possibilities and future directions from the technological point of view. This shift from the conventional centralised data processing, moving the monitoring task direct to its source will be the main challenge for the future generations of the WINS (wireless intelligent network sensors).

10 ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the European Union Sixth Framework programme for Research and Technological Development. This paper summarises work performed as part of FP6 Integrated Project IP017498 DYNAMITE Dynamic Decisions in Maintenance.

11 REFERENCES

[1] http://www.ewh.ieee.org/tc/sensors/Field_of_Interest.htm - IEEE sensors council; last accessed December 2005

[2] Bojko, T. (2005), Wireless sensors network for vibration measurements, Pomiary Automatyka Kontrola, n 5, p 12-15

[3] Simonen, J.T. Andringa, M.M.; Grizzle, K.M.; Wood, S.L.; Neikirk, D.P., (2004), Wireless sensors for monitoring corrosion in reinforced concrete members, Proc SPIE, v 5391, n 1, p 587-96

[4] Ihler, E, Zaglauer, H W.; Herold-Schmidt, U; Dittrich, K W.; Wiesbeck, W, (2000), Integrated wireless piezoelectric sensors, Proc SPIE, v 3991, p 44-51

[5] Chuang J, Thomson, DJ (2005) Noise reduction in RF cavity wireless strain sensors, Proc SPIE, v 5768, n 1, p 344-53

[6] Reindl, L. Shrena, I.; Kenshil, S.; Peter, R. (2003) Wireless measurement of temperature using surface acoustic waves sensors, ProcIEEE Int Frequency Control Symposium (IEE Cat. No.03CH37409), p 935-41

[7] Han, T, Shi, W (2001) Surface acoustic wave devices based wireless measurement platform for sensors, Proc SPIE, v 4601, p 14-19

[8] Kohlstrand, K.M., Danowski, C.; Schmadel, I.; Arms, S. (2003) Mind the gap: using wireless sensors to measure gaps efficiently, Sensors, v 20, n 10, p 14, 16-20

[9] McGrath, M., Sabouni, R.; Pham, A. (2004), Development of nano-based resonator gas sensors for wireless sensing systems; Proc SPIE, v 5593, Nanosensing Materials and Devices, p 62-72

[10] Clayton, E.H.; Bong-Hwan Koh; Guoliang Xing; Chien-Liang Fok; Dyke, S.J.; Chenyang Lu (2005) Damage detection and correlation-based localization using wireless mote sensors, IEEE Int Symp Intelligent Control and 13th Mediterranean Conference on Control and Automation (IEEE Cat. No. 05CH37647), pt. 1, p 304-9 Vol. 1

[11] Atashbar, M.Z., Bazuin, B.J.; Krishnamurthy, S. (2003) Design and simulation of SAW sensors for wireless sensing, Proc IEEE Sensors 2003 (IEEE Cat. No.03CH37498), pt. 1, p 584-9 Vol.1

[12] Lamprinos, I.E.; Prentza, A.; Nokas, G.; Dermatas, E.; Tsoukalis, A.; Koutsouris, D. (2004) Design of wireless network of sensors for continuous monitoring of vital biosignals, Proc IEE Eurowearable '03, p 79-85

[13] Wang, Ning; Zhang, Naiqian; Wang, Maohua (2006) Wireless sensors in agriculture and food industry - Recent development and future perspective, Computers and Electronics in Agriculture, v 50, n 1, p 1-14

[14] Moschler, W W.; Hanson, G, (2004) Wireless microwave based moisture sensors for hardwood drying kilns, Proc AIChE Annual Meeting, p 1707-1708

[15] Kiefer, K.F., Swanson, B.; Krug, E.; Ajupova, G.; Walter, P.L. (2003) Wireless sensors applied to modal analysis, Sound and Vibration, v 37, n 11, p 10-17

[16] Asada, G., Dong, T. S., Lin, F. N., Pottie, G., Kaiser, W. J., Marcy, H. O. (1998). Wireless integrated network sensors: Low power systems on a chip. European Solid State Circuits Conference, The Hague, Netherlands. http://citeseer.ist.psu.edu/278712.html

[17] Boltryk PJ, Harris CJ, White NM, (2005), Intelligent sensors - a generic software approach, Journal of Physics: Conference Series 15 Sensors & their Applications XIII (2005) 155–160, Institute of Physics Publishing

[18] http://www.sensornetworks.net.au/intsens.html; last accessed December 2005

[19] http://www.nsti.org/Nanotech2003/showabstract.html?absno=601; last accessed December 2005

[20] http://www.automationworld.com/articles/Features/1132.html; last accessed December 2005

[21] Esteban J, Starr A G, Willetts R, Hannah P, Bryanston-Cross P, 2005, A review of data fusion models and architectures: towards engineering guidelines, *Neural Computing & Applications* Springer-Verlag London Ltd, ISSN: 0941-0643 (Paper) 1433-3058 (Online), Volume 14, Number 4, pp273-281

[22] http://www.wirelesssensingconnector.com/; last accessed December 2005

[23] Gregory J. Pottie and Loren P. Clare, (1998), Wireless Integrated Network Sensors: Towards Low Cost and Robust Self-Organizing Security Networks, Proc. SPIE Sensors, C3I, Information, and Training Technologies for Law Enforcement Vol. 3577, p. 86-95

[24] http://www.isr.umd.edu/ISR/publications/newsletter/ssfa00/ShammaDARPA.html; last accessed December 2005

[25] http://www.sensorsmag.com/articles/0402/40/; last accessed December 2005

DECISION SUPPORT SYSTEMS IN PLASTIC RECYCLING PLANTS

C Hans[a], M Schnatmeyer[a] KD Thoben[a] I Sbarski[b] & John Bishop[b]

[a] *Bremen Institute of Industrial Technology and Applied Work Science at the University of Bremen (BIBA), Hochschulring 20, 28359 Bremen, Germany.*

[b] *Industrial Research Institute Swinburne, PO Box 218, Hawthorn 3122, Australia.*

Abstract: For higher recycling quotas of EOL goods, the recycling industry needs economic strategies for optimized reverse supply chains and recycling processes. Especially in the EOL area with a high degree of mixed materials, appropriate systems have to be used for supporting the recycler with decisions during the recycling of products. This paper aims to give an overview on relevant approaches which can help recycling companies to deal with the challenges of this highly dynamic environment. Furthermore, available tools and their application in the plastic recycling area will be presented.

Key Words: Automation, Communication systems, Complex Systems, Computer-aided work, Computer-integrated manufacturing, Decision support systems, Interfaces, Manufacturing processes, Tracking applications.

1 INTRODUCTION

Plastic Recycling Plants are traditionally related on the Small and Medium Sized Enterprise (SME) sector. Here decisions are usually more related on personal knowledge of the staff than on automated decision support systems. This causes insufficient processes because knowledge gained from single recycling processes cannot be reused for new production lots on customer demands. Another important issue, which makes this sector quite complex, covers the uncertainties related to the emergence of recycling material. Although there are predictable scenarios where for instance residuals from production are continuously delivered to recycling, processes are in more or less stable quantities. However, there are many other scenarios where the available material appears in unpredictable quantities. Furthermore the quality and composition of the incoming material can vary dramatically. In any case the recycling facilities are expected to fulfil the demand of their customers.

1.1 Sector description

In general the sector of recycling and especially the recycling of plastics is indicated by various uncertainties. The companies active in this field have to deal with various quantities which are delivered for recycling as well as volatile demands from the customers. These factors make reliable planning strategies very difficult. Another typical problem in the area of plastic recycling is the quality of incoming materials. The deterioration of recycled materials properties as compared to the virgin plastics can be caused by contamination, environmental degradation of the plastics in the course of their application, and thermal and mechanical loads during the recycling process. Examples for such quality indicators are thermal, mechanical or rheological properties of materials such as melting or glass transition temperature, elongation, yield strength, flexural strength, melt flow index etc. [1].

As it is seen from the Table 1 there might be a substantial difference between the properties of virgin and recycled materials that can effect their processing and as a result properties of end-use products. Thus today companies in the recycling sectors aim to react to different situations as flexible as possible. In many cases the proper reaction depends on the experiences of few people. In this context, a decision support system would be desirable which supports the decision maker and furthermore offers new potential for optimisation of the often complex recycling processes. At the end, such systems could help to make the whole process more transparent and thus traceable.

Table 1

Some properties of typical recycled polyethylene terephthalate (PET) compared with virgin PET

Properties	Virgin PET	Recycled PET
Melting point, °C	241	252
Recrystallisation temperature	176	193
Tensile modulus, MPa	2321	2367
Creep deformation in 0.5 min (at 100 °C and 0.05 MPa tensile load), %	2.1	3.2
Viscosity(temperature 270 °C, shear rate 25 1/s), Pa s	100.7	37.4

1.2 Decision support systems - State-of-the-Art

Depending on the application field there are different approaches for decision support systems (DSS). Each of them comes along with its own strengths and weaknesses and thus is applicable only for certain areas. However, keeping in mind the diversity of the decision problem, it is absolutely necessary to be aware of the main problem areas related to recycling. In this paper three problem areas in the recycling process are considered as the main ones:

1. Uncertainties regarding material quantities and qualities
2. Material treatment
3. Production, warehousing and logistics

For all of these areas particular approaches can be identified supporting the decision maker. The most important ones are listed below together with a reference to the problem area:

- Simulation (1, 3)
- Formal (mathematical) approaches (3)
- Expert Systems, Neural Networks (2, 3)
- Benchmarking (2, 3)
- Specific methodologies (2)

The list above shows that none of the methods is able to cover all of the problem domains. Thus a combination of several is desirable. Although all of the concepts mentioned are well defined and have been used for decision support in various areas, none of them are available in the form of user friendly tools. In particular, formal approaches like mathematical programming (e.g. integer / mix-integer programming) are used for problems related to production warehousing and logistics. However, significant knowledge is required in order to apply them which makes them unusable for domain experts as decision makers usually are. Methods from artificial intelligence, like neural networks or expert systems, usually suffer in this context from the necessity to train them or to specify complex knowledge bases. Their application is therefore often limited to certain scenarios. An extension requires significant effort. Benchmarking approaches require a significant number of cases / scenarios in order to be useful. The main problem in this context covers the willingness to deliver internal data.

Thus the most promising candidates for effective decision support systems in the field of plastics recycling are simulation-based approaches as well as certain methodologies which have been specified for particular problem domains together with accompanying tools in order to simplify their application for domain experts. In addition to the requirements regarding applicability as well as user-friendliness previously mentioned, an important issue in this context covers the integration of these approaches / environments into the value creation chain. Existing solutions will be shortly presented in the following chapter.

1.3 What can innovative technology offer?

An interesting question in this context covers the application of innovative concepts – like the concept of Product Embedded Information Devices (PEID) - which is considered in the PROMISE project (see section 3).

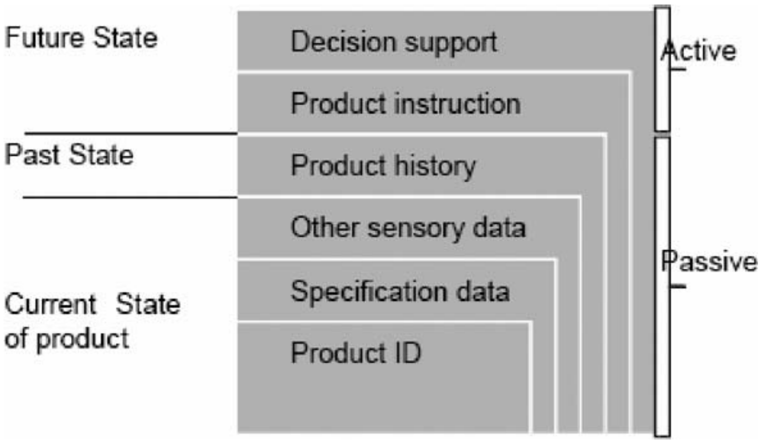


Figure 1. Product Information Model Enabled by Auto-ID [2].

In fact such technologies offer considerable potential to be exploited for decision support. As shown in Figure 1, the active behaviour of product entities could be utilised in order to develop completely new environments or enhance existing ones by making them more responsive due to the pervasive availability of product related information anytime and nearly everywhere. In particular available model-based solutions (like simulation) can benefit from such technologies because their main deficiency is in the timeliness of the information kept by the underlying model. Appropriate concepts exploiting the available information can reduce the costs and time for solution finding dramatically.

2 DSS INTEGRATION

With respect to the integration of the simulation-based DSS – ONE (see section 3) into the current scenario, the main goal is to provide real-time data for incoming materials and goods on the basis of the required specifications for PEIDs and RFIDs within the scenario. Real-time data will be provided for all of the relevant process steps within the scenario for the support of back-end systems such as Warehouse Management Systems (WMS) and Production Planning Systems (PPS). Afterwards the planning system will be able to simulate basic processes like purchasing and sales planning taking into account the availability of materials, maintenance costs, etc. on the basis of real-time in-formation provided by PEIDs.

All the real-time data information logistics which are offered by the PROMISE solution will be handled by the middleware. That means, the PROMISE middle-ware needs to be capable of managing the data flows among the recycling machines, transport systems, and control systems. Furthermore, it must provide the links between the software embedded in the PEIDs, PEID readers and back-end systems and should be based on platforms used in PROMISE and is compatible with other application scenario implementations.

In order to handle the potentially vast amount of product information generated within the scenario, the middleware must furthermore provide suitable filtering mechanisms for the data flow between the system components and especially toward back-end systems. To that end, it should include mechanisms such as plausibility checking, duplicate data avoidance, data persistence monitoring, etc. In that context, depending on the design and specification of both the middleware and the DSS, it is possible for the middleware to incorporate DSS functionality where suitable. Last, but not least, the middleware is expected to integrate suitable security mechanisms for protection of personal and process relevant data.

For the integration of the ONE tool with the PROMISE environment a pure data-based interface should be considered (at first) for the first step. Although more or less manual, this approach would allow the connection for less costs and efforts than for a more direct coupling which can be looked at a second step. Thus the exchange of data will be conducted via file creation, transmission and import into the ONE system. Especially for file creation internal mechanisms provided by the PROMISE platform will be applied in order to generate the right format and content. The import process is shown by the following screen shot (Figure 4). For each variable the system allows the specification of a constant value, according to a certain distribution function or in the form of a file where the data will be taken from. The example below shows the specification of a certain user demand.

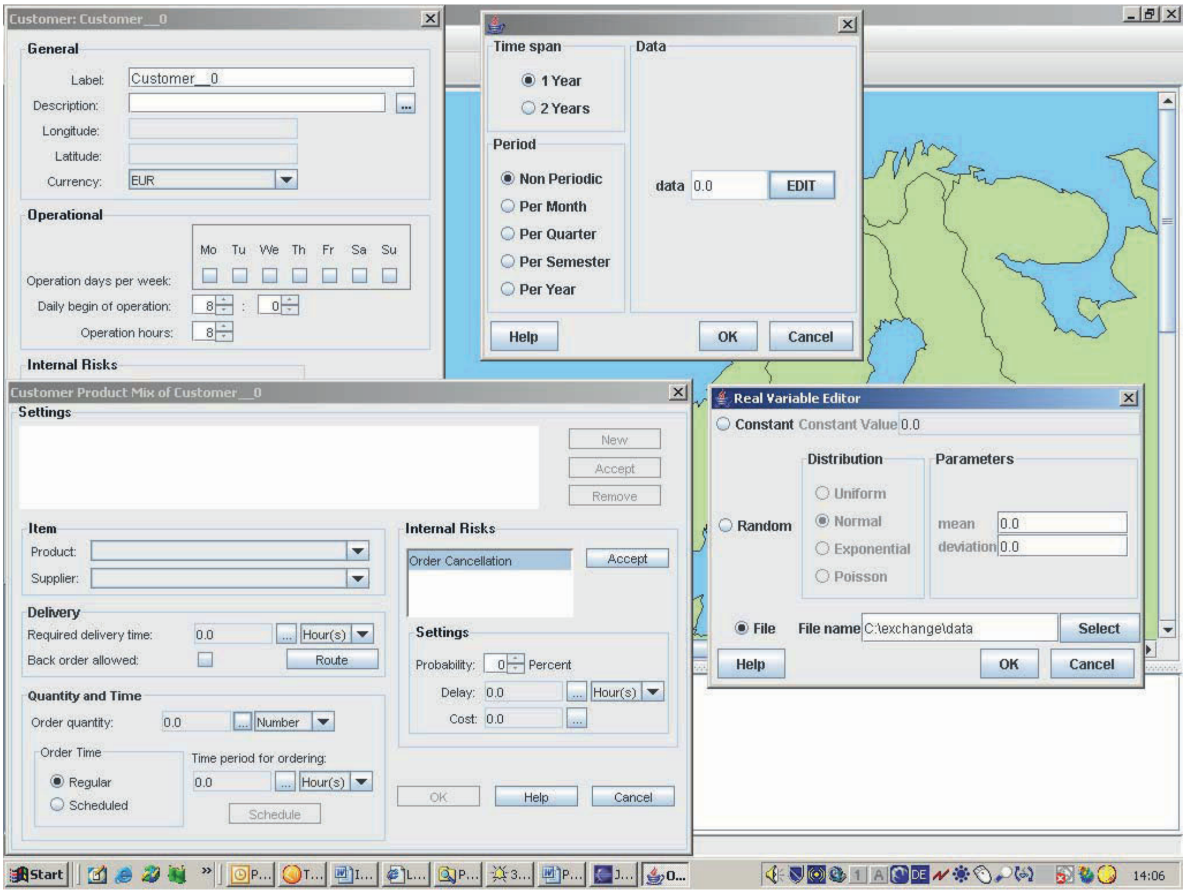


Figure 2. Data import in ONE.

After the import of the data representing customer demand for a certain period in time the DSS can support purchase planning as well as sales in determining available materials and sales planning. In addition, logistics solutions (distribution networks, etc.) can be planned and evaluated before being implemented in reality with the simulation component. Thanks to this component it is also possible to consider uncertainties regarding customer demand or the quantity of material available for recycling. Additionally, process related risks can be investigated, e.g. regarding the impact of delays of deliveries due to accidents or the technical breakdowns during the recycling process.

3 RESEARCH BACKGROUND

The described integration concept is based on different research projects, which will be shortly described in this chapter.

3.1 Product Lifecycle Management and Information using Smart Embedded Systems (PROMISE)

PROMISE is an international research project in the IMS (Intelligent Manufacturing Systems) program with industry and academia consortia in Australia, Europe, Japan and the United States (see Figure 3). The project concerns the whole information flow from Design, Production, Use-Service-Maintenance or MOL (Middle-of-Life) and Retirement or, as most commonly is called, EOL (End-of-Life). The European part (funded by the European Commission) will develop Information Models, Architecture and Implementation concepts for the PEID technology. This includes also models for decision making.

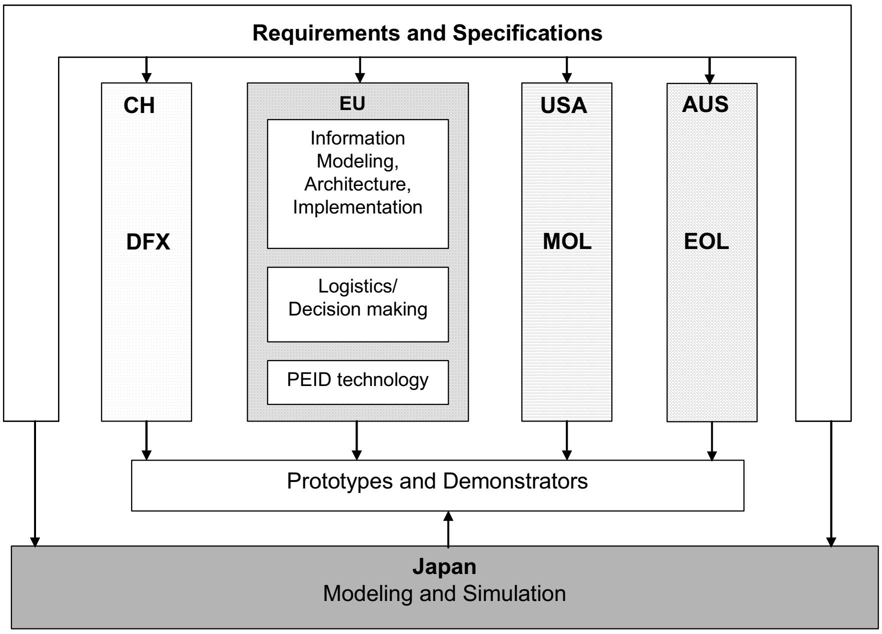


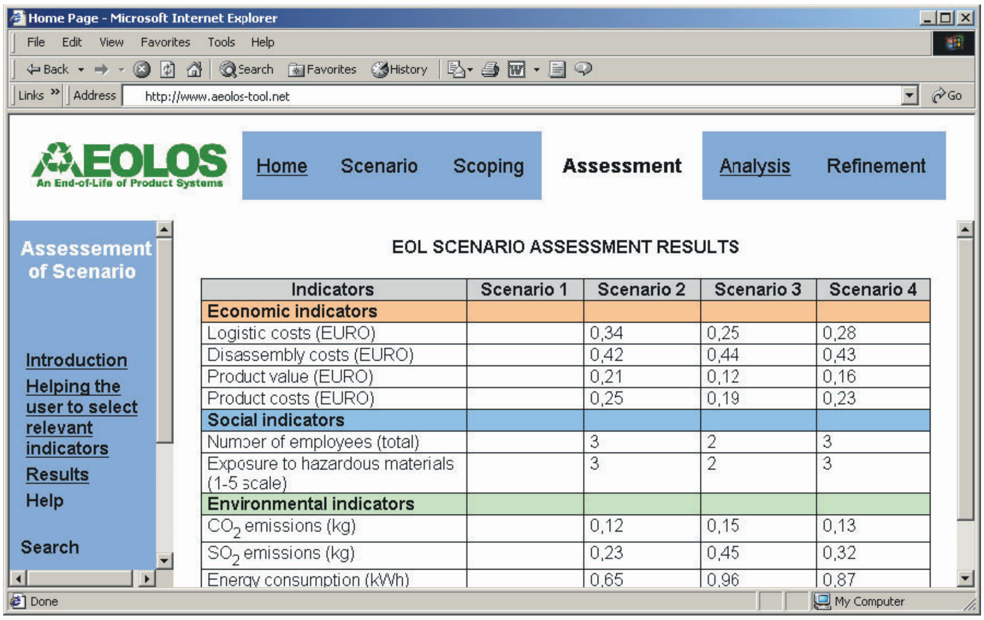
Figure 3. IMS PROMISE regional cooperation [3].

3.2 An End-of-Life of Product Systems (AEOLOS)

The methodology and toolset developed in AEOLOS will provide support to decision-makers both within the public sector and in industry. Stakeholders addressed by the AEOLOS are:

- Society"
    - as a heterogeneous group, and including future generations
    - as particular communities or interest groups
    - as represented by government at all levels
- Manufacturers
- Remanufactures
- Reclaimers
- Recyclers
- Incinerator operators
- Landfill operators
- Title owners of product and
- Distributors.

Decision-makers in all of these areas formulating policy at local authority, regional, national and EU levels will be supported with the means by which to determine, analyse and evaluate the most appropriate EOL treatment option from an environmental, economic and social perspective for their geographical areas of concern. Similarly, resource recovery/recycling companies and original equipment manufacturers (OEMs) will be provided with decision support as to the choice of EOL treatment options, based on a set of environ-mental, economic and social criteria. The solution is progressing with a methodology and a web-based toolset supporting it's application (see Figure 4 below).



EOL SCENARIO ASSESSMENT RESULTS

Indicators	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Economic indicators				
Logistic costs (EURO)		0,34	0,25	0,28
Disassembly costs (EURO)		0,42	0,44	0,43
Product value (EURO)		0,21	0,12	0,16
Product costs (EURO)		0,25	0,19	0,23
Social indicators				
Number of employees (total)		3	2	3
Exposure to hazardous materials (1-5 scale)		3	2	3
Environmental indicators				
CO_2 emissions (kg)		0,12	0,15	0,13
SO_2 emissions (kg)		0,23	0,45	0,32
Energy consumption (kWh)		0,65	0,96	0,87

Figure 4. AEOLOS web-based toolset [4].

3.3 Optimisation methodologies for Networked Enterprises (ONE)

The project ONE (Optimisation methodologies for Networked Enterprises) developed an integrated decision support tool for the evaluation and optimisation of existing as well as the planning of new supply chain networks. In this context not only monetary aspects were considered but also environmental and social impacts, along with certain network specifications. Furthermore, the tool supports risk assessment which means the analysis of the robustness of supply chain networks against unpredictable events such as accidents, bankruptcies or natural disasters. As another important issue, the tool is able to deal with uncertainties like volatile customer demand. The ONE-System encompasses the following components:

- A network modelling module providing user friendly support for graphical, interactive development of supply chain networks.
- A statistical data miner allowing the analysis of existing data material and the extraction of information (e.g. in the form of specific probability distributions for customer demand) in order to make network descriptions more realistic.
- A set of optimisation tools each of which supports the optimisation of certain aspects within a network model.
- In order to prove the impacts of changes or network configurations proposed by the optimisation module, ONE includes a simulation engine for the emulation of a network model.

Especially the consideration of uncertainties as well as risk together with the combination of optimisation methodologies and simulation makes the ONE system an interesting approach for decision support while dealing with today's highly dynamic supply chain network planning and optimisation. The combination of these functionalities make ONE a prime candidate for integration with new technologies like the one which is proposed by the PROMISE project.

4 CONCLUSIONS

The paper gives an overview about the specific requirements in the field of plastics recycling. Furthermore, different approaches for decision support within this area were discussed. It was shown that different concepts come along with their own strengths and weaknesses. Due to the requirements regarding applicability, user-friendliness and integration into the value creation process most of these concepts are not suitable. On the other hand, even promising environments – namely simulation as well as specific methodologies and the accompanying toolsets - are not able to cover all end-users needs on their own. Thus a combination of both concepts would be desirable. Furthermore it was detected that the most challenging requirement clearly concerns the integration of DSS tools into "real life". In this context the paper gave an overview about a possible approach demonstrated along the simulation-based DSS tool ONE. In addition, the potential benefits coming along with innovative concepts like the one which is considered by the PROMISE project were addressed. In this context major improvements regarding the accuracy of the results from available DSS solutions can be expected due to the improvement of the underlying model (information).

The PEID concept perfectly fits to the concept of GRID computing for the globally distributed topology comprising product entities and the associated information. Formerly more utilized in the field of high performance computing, GRID computing is nowadays investigated regarding it's suitability for other application fields, e.g. distributed enterprise networks. However, the integration of PEID and GRIDs can be an interesting approach, especially for the development of a new generation of decision support systems which are able to consider global environments for solution finding.

5 REFERENCES

[1] Bishop, J., I. Sbarski and E. Kosior (2004) Reuse of Post Consumer Paper and Plastic Materials. Proceedings International IMS Forum 2004, S. 1488 – 1495. Biassono (Milano).

[2] Mcfarlane, D. (2003) The impact of Product Identity on Industrial Control Part 1: "See More, Do More…". White Paper CAM-AUTOID-WH-012, Auto-ID Center, MIT. Available at www.autoidlabs.org.

[3] PROMISE (2005) Product Lifecycle Management and Information using Smart Embedded Systems. http://www.promise.no (07/11/2005).

[4] Goggin, K. and Schnatmeyer, M. (2002) AEOLOS - An-End-Of-Life of Product Systems. Proceedings EUROSUSTAIN 2002 Implementing the Integrated Product Policy, CD-ROM. Rhodes, Greece.

6 ACKNOWLEDGMENTS

The authors wish to acknowledge the European Commission, AusIndustry in Australia and the PROMISE consortium for their support.

CONDITION BASED RISK MANAGEMENT (CBRM) – ENABLING ASSET CONDITION INFORMATION TO BE CENTRAL TO CORPORATE DECISION MAKING

D Hughes[1] G Dennis[1], J Walker[1] and C Williamson[2]

[1]EA Technology – UK [2]Thew & Mc Cann Pty Ltd - Australia

Abstract: CBRM is a process developed by EA Technology in conjunction with several major electricity companies [1-4] to assist with the tasks of defining, justifying and subsequently targeting spending to achieve defined levels of performance. The process has already delivered major benefits. If fully embraced, it can become the basis for effective asset management programmes that address the critical issues of renewing networks and maintaining reliability within an ever more challenging regulatory and financial environment

Key Words: CBRM, asset condition

1 THE CBRM PROCESS

The essence of CBRM is the provision of a structured framework that enables available engineering knowledge and experience of network assets to be readily and transparently linked to corporate decision making. Application of the process results in the fundamentally important decisions, relating to both operational and capital spending, being directly linked to asset condition and future performance.

One of the great strengths of distribution network owners and operators is the wealth of available asset specific and generic knowledge and experience relating condition, performance, degradation and failure mechanisms, operating context, maintenance regimes etc. CBRM enables this to be effectively utilised to deliver corporate objectives.

To understand the mechanics of CBRM it is useful to define the process by a number of sequential steps as listed below. A more complete understanding and appreciation of the potential of the process can then be obtained from discussion of its application and examples of the output in later sections.

1. **Define asset condition**. Derive 'heath indices' for individual assets and build health index profiles for asset groups. Health Indices on a scale of 0-10, 0 indicating the best condition, 10 the worst.
2. **Link current condition to performance**. Calibrate the health index against relative probability of failure (POF). Match the health index profile with current failure rate to determine health index/POF relationship.
3. **Estimate future condition and performance.** Use knowledge of degradation processes to 'age' health indices, ageing rates dependent on initial health index and operating conditions. Calculate future failure rates from aged health index profiles and previously defined health index/POF relationship.
4. **Evaluate potential interventions in terms of POF and failure rates**. Factor in the effect of potential replacement, refurbishment or changes to maintenance regimes, modify future health index profiles and recalculate future failure rates.
5. **Define and weight consequences of failure (COF)**. Construct and populate a consistent framework to evaluate consequences in significant categories, safety, network performance, environmental, financial etc.
6. **Build a risk model**. Combine POF and COF for asset groups to quantify risk. Total risk can be separated into previously defined categories. Total risk and risk within each category related to tangible quantities, £, CMLs, frequency of fatalities or serious injuries etc.
7. **Evaluate potential interventions in terms of risk**. Factor in the effect of potential replacement, refurbishment or changes to maintenance regimes, recalculate POF and COF and quantify risk reductions
8. **Review and refine information and process**. Learn from applying the process, identify opportunities to improve asset information and refine models and algorithms. Define and progressively build an improved asset information framework.

2 APPLICATION OF CBRM

2.1 Generating short term results

It is important to emphasis that CBRM has been developed as a result of working with individual electricity companies to achieve specific, short term objectives. In particular to assist companies to prepare for regulatory submissions. To define asset condition and create the framework for evaluation of potential investment plans within a relatively short and clearly defined timescale necessitated an approach that did not involve major information gathering. Therefore the initial applications were based on the principle of maximum utilisation of existing information.

Within every electricity company we have worked with (on CBRM projects or in other areas) we have encountered extensive knowledge and experience of almost all asset classes.

In addition to the specific internal information, there is also extensive knowledge and experience of degradation and failure processes for most asset groups. There is a good understanding of the critical issues that relate environment, duty, maintenance etc to the condition and performance of assets and huge body of generic (type specific) experience within individual asset groups.

CBRM is about using all this engineering knowledge and experience to maximum effect, creating a consistent framework to utilise this to define condition and link to current and future performance. It is about putting engineering knowledge and experience right at the centre of asset management decision making.

Our experience is that for major asset groups, including substation assets, overhead lines and cables from 11kV upward, health indices that provide a reasonable definition of condition and can be justifiably related to POF, are viable. Clearly, the nature of information available is quite variable and it is accepted that these initial health indices are not 'perfect', but based on the extent of engineering knowledge available they are deemed to be a credible definition of condition and performance.

A very important aspect of CBRM is that it requires a critical assessment of available information and identifies omissions that, if addressed, would greatly increase the confidence in the health index. It also clearly illustrates how information could be defined and collected in a manner that enhances future asset management processes. This creates opportunities to systematically improve asset information as discussed in later sections.

Having completed projects with many companies we have realised the value of working within a common framework relating health indices to POF. This had led to use of standard curves that are fitted to specific health index profiles and failure rates for each population. This enables direct comparison between different assets or similar assets in different locations and enables health index values to relate to consistent definitions of condition and performance. In essence, we now 'calibrate' each set of results against a common definition of condition and performance. By cross referencing results for similar assets in different locations with different performance levels we can greatly increase the confidence in the results.

A feature of the health index, POF and ageing methodologies used in CBRM is that at all stages they are related back to physical condition and degradation and failure processes and assessed in light of the practical experience of the assets. In this way there are a number of 'reality checks' that can be applied to the results and the predictions of future performance.

In the same way, the assessment of consequences and building of the risk model is based on practical experience and calibrated against tangible values. In most cases the actual total risk in a specific category (safety, financial, network performance) can be defined in real terms. Thus the final 'risk figure' for an asset group and the benefit in terms of risk reduction resulting from a particular investment programme can also defined using the same real values.

The CBRM process has been successfully applied to the full range of assets that make up distribution and transmission networks. The value obtainable in the short term is illustrated by example in the following section.

2.2 Examples of CBRM output

Health indices, POF and failure rates. The first output and the foundation for the rest of the process is the definition of asset condition in the form of a health index derived for individual assets. These are built into a health index profile, a distribution of health indices, for a population.

As described earlier, considerable thought has gone into the relationship between the health index and POF. An exponential relationship was initially identified as a reasonable description and subsequently this has been used to derive and calibrate health indices. The form of the basic relationship is shown schematically figure 1.

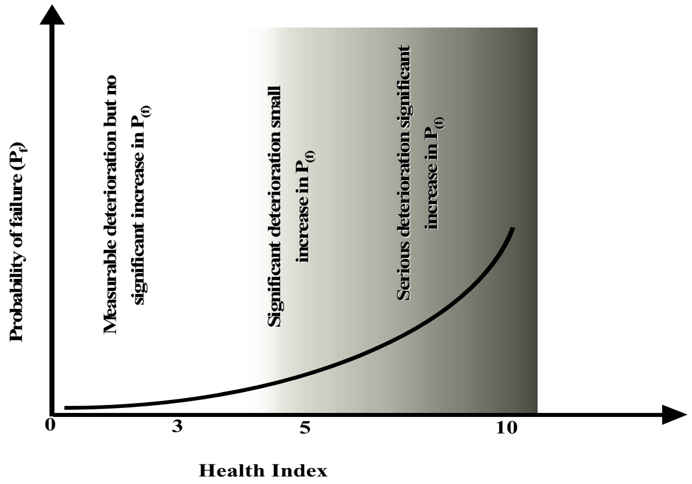


Figure 1

By adopting a consistent basis for the relationship between health index and POF, a health index profile gives an immediate appreciation of the condition of assets in a group and an understanding of the implication for future performance.

A health index derived for a population of HV OHLs based on a combination of condition and design information is shown in the profile in figure 2. Ageing this profile and factoring in the improvements in performance that would be obtained by rebuilding lines enables an estimate of the overall improvement in performance for different levels of investment, as summarised in tables 1 and 2.

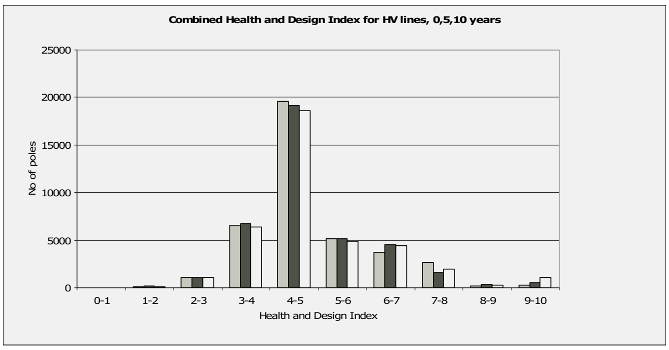


Figure 2, Combined condition and design index for HV OHLS

	Condition and design	3rd Party	Total
Current rate	0.064	0.028	0.092
Year 5	0.072	0.028	0.100
Year 10	0.085	0.028	0.113

Table 1: Fault rates (per km) assuming no rebuild

% of lines rebuilt over 10 years	Condition and design	3rd Party	Total
2.5	0.063	0.028	0.091
5.0	0.059	0.028	0.087
8.0	0.048	0.028	0.076
20	0.035	0.028	0.063
33	0.028	0.028	0.056

Table 2: Fault rates in year 10 for different levels of rebuild

Another example (figures 3&4) shows the health index profiles for a population of large transformers. In this case, different ageing algorithms are applied to transformers based on loading and environment.

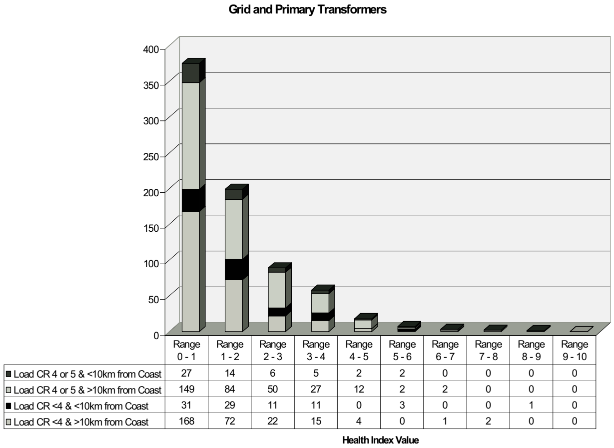


Figure 3, Current profile for population of grid and primary transformers

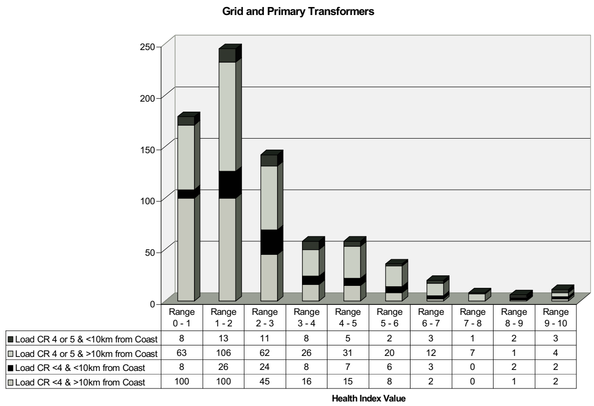


Figure 4, Profile in year10, assuming no significant intervention

Applying the calculated HI/POF relationship to the Year 10 profile indicates a 3 fold increase in rate of failure. Factoring in interventions reveals that to maintain the current (acceptable) failure rate would require 92 (out of 850) transformers to be replaced over the 10years. By considering the relevance and effect of refurbishment/remedial measures on an individual transformer basis and factoring these in, the same overall result can be obtained by a significantly cheaper option involving replacement of 45 transformers and refurbishment/remedial work on 70.

Risk calculations. The initial applications used health indices and POF calculations to define investment requirements to achieve a specified level of failure rate. More recently risk models (combining COF with POF) have been populated to provide a more complete means of optimising investment.

Risk is evaluated in 4 basic categories (financial, safety, network performance, and environmental). In each category the consequence of failure (the range and the distribution) are determined by consultation with experienced engineers and is calibrated against real values wherever possible, i.e. the number of CMLs resulting from failures of the asset group.

The relative weighting between each category is determined by relating each consequence to a common framework, value in monetary terms, e.g. the monetary value of a CML. This enables the overall risk for any scenario (with or without an investment package) to be calculated as a single number (of risk units, RUs). The total number can then be split down into the 4 categories and the effect (reduction in risk) achieved by different investments packages can be compared, as illustrated by the example in table 3 below.

A risk summary for a population of fluid filled cables.

Base Case, at the end of next regulatory period with no significant investment.

Option 1, including a programme to replace short sections, install additional stop joints and remake terminations at a cost of £12m.

Option 2, Option 1 with increased cable replacement at a cost of £16.5m

Table 3, Summary of risk for fluid filled cables

Risk Category	Base Case	Option 1	Option 2
Network performance	1,893	759	670
Safety	584	352	292
Financial	18,510	6,308	5,552
Environmental	21,653	6,980	6,114
Total	**42,640**	**14,399**	**12,628**
Benefit		28,241	30,012
Cost		£12m	£16.5m
Cost Benefit		424£/RU	549£/RU

1RU ≡ 1500CML,   1RU ≡ £70 OPEX,

50,000RU ≡ 1 fatality, 1000litres of oil ≡ 140RU

This risk model is still in the development stage and needs some refinement, nevertheless the results we are achieving provide an effective means of defining a complex and rather abstract quantity (risk) in a manner that can be related to tangible values. Encouragingly where the results have been presented and discussed with asset managers there has been a high level of satisfaction, both in terms of the relative values between asset groups and the overall presentation of risk, in a manner that can be directly related to engineering knowledge.

3 SHAPING THE FUTURE

3.1 Building CBRM into business as usual

The initial impetus to develop CBRM came from the need to define and justify spending plans in regulatory submissions. Great success has been achieved but it has been a labour intensive process to provide a one off snapshot of asset condition and risk. The potential value of building CBRM into everyday activities such that asset managers routinely have access to well defined asset condition information that can be used to actively manage asset is clear. Indeed the ability to target spending (both OPEX and CAPEX) to achieve defined benefits will be essential to maintain acceptable reliability while working within strict financial restraints.

'Built in' CBRM would give asset managers the ability to direct resources to achieve specific performance objectives, something that is absolutely essential in the increasingly demanding regulatory and financial environment.

3.2 Improved asset information – the key to future success

The key to implementing CBRM is access to appropriate asset information and therefore the key to building CBRM into business as usual is the creation of an information framework that gives asset managers access to appropriate information routinely. Improved asset information to enhance asset management decision making is seen as the key to success. Discussion of improved asset information usually leads to thoughts of huge IT projects, which involve mammoth cost and have very distant deliverables (probably the last thing electricity companies need at present). Contrary to this, the experience of implementing CBRM has revealed a real opportunity to achieve this objective (routinely available, quality information) relatively painlessly, incurring only modest initial costs with overall cost savings in the medium term!

During the development and application of CBRM, a real understanding of the nature and quality of information currently held by electricity companies has been obtained. This has highlighted the present deficiencies and provided an insight into the means of progressively improving the quality, relevance and consistency of asset information in an economic and practical manner.

A central feature of the CBRM process is the critical review of the information current available. This inevitably leads to identification of additional or alternative information that could be particularly useful in improving the ability to define condition and future performance.

Furthermore, activities involving extracting data from various sources and combining it to give a numerical value of health index, has demonstrated the difficulty of working with data in the variety of different forms currently available. The potential value of adopting a system that defines the results of all inspection or tests by a relatively simple numeric condition rating system then becomes clear.

The CBRM process, and some of the work to optimise maintenance that preceded it, also demonstrated the significance and importance of maximising opportunities to obtain critical condition information, in particular to integrate the information collection across the inspection and maintenance activities. This leads to significant opportunities to optimise the whole management process and generates significant OPEX savings, better targeted inspection leading to less invasive maintenance and the collection and storage of less (but more useful) data.

Another feature of the CBRM process is the principle of progressive improvement. The aim has always been to produce the best result possible with the currently available information and then define the means to progressively improve the output by improving the information. Improvements in information are obtained not with large expensive condition assessment activities, but by identifying potential useful information that can be obtained during existing, routine inspection and maintenance activities.

This is perhaps the most critical lesson. Distribution assets are visited frequently. By careful consideration of the information that will enhance decision making processes it is possible to define inspection protocols, using simple numeric condition ratings that will rapidly lead to a data resource that can routinely enable much more effective decision making. By targeting the information to meet specific requirements, the volume of data and the whole inspection and maintenance regime can be rationalised.

All this should be possible using existing IT infra structure. Creating an improved information framework is an engineering issue not a major IT project!

Applying the principles discussed above therefore provides electricity companies with a significant opportunity to inexpensively create an information framework that will enable targeted spending to deliver specific network performance at minimum cost. This is critical to future success. For UK companies this will enable delivery of DPCR4 targets and be excellent preparation for DPCR5!

Much of the development of CBRM has been in conjunction with UK electricity distribution network operators, initially in preparation for their last price review (2004). Subsequently several companies have embraced the principles of CBRM and are currently building them into their routine asset management activities both to provide the cornerstone of their next price review submission.

In recent months significant CBRM projects have been undertaken with major transmission companies in locations that include, China, Far East and Middle East. It appears that CBRM is an effective asset management tool in any language!

4 CONCLUSIONS

CBRM provides an effective means of linking the extensive engineering knowledge and experience of network assets, available in most companies, to corporate decision making.

Implementation has demonstrated it can deliver significant short term benefits.

The CBRM experience offers the opportunity to build an improved asset information base that would greatly enhance decision making capability. It would enable future spending to be defined, justified and targeted to deliver specific levels of network performance at minimum cost.

CBRM is a vital component of successful asset management in the ever more demanding regulatory and financial climate.

5 ACKNOWLEDGEMENTS

The authors are indebted to the many colleagues at EA Technology who have contributed significantly to the development of CBRM and to the numerous ‘partners’ in the participating electricity companies. CBRM only works if there is effective collaboration between all those who share knowledge and experience of the assets.

6 REFERENCES

[1] E Hamilton and D Hughes. ‘CBRM, the practical application of intelligent asset management to United Utilities distribution network’, Distribution Europe 2004, April 2004.

[2] D Hughes. ‘The use of health indices to determine end of life and remnant life of distribution assets’, CIRED, May 2003.

[3] T H Jartouson, B Jesus, D Hughes and M Godfrey. ‘Development of health indices for asset condition assessment’. IEEE, Transmission and Distribution, Dallas, USA, September 2003.

[4] T Richards and D Hughes, ‘The use of condition information for strategic risk management’, Euro TechCon Birmingham, November 2002.

THE SOCIETY FOR MAINTENANCE & RELIABILITY PROFESSIONALS (SMRP)

C Armbruster [a] & T Byerley [b]

[a] SMRP Chair, www.smrp.org, / Corporate Reliability Engineer Manager,The J. M. Smucker Company

[b] SMRP Vice Chair, www.smrp.org / Director, The Maintenance & Reliability Center, The University of Tennessee

Abstract: The Society for Maintenance & Reliability Professionals (SMRP) is an organization formed by practitioners, for practitioners. SMRP is committed to and an advocate for excellence in Maintenance and Reliability professionals.

SMRP is incorporated as an independent, not-for-profit professional society. As such, SMRP Bylaws state the Society is organized and operated exclusively for educational, scientific and charitable purposes.

The Society has grown to international status with its vision to become *the* global organization best known for creating value through improved physical asset management

Key Words:maintenance practitioners,

1 INITIATIVES FOR MAINTENANCE PRACTITIONERS

1.1 Our History

SMRP was formed and chartered in 1992 to network those in the maintenance & reliability profession. SMRP continues that dedication, instilling excellence in Maintenance and Reliability in all types of manufacturing, service, and educational organizations.

1.2 Our Mission

SMRP will be the global leader:

- Promoting and supporting Maintenance and Reliability education for people, production and quality processes to improve the work environment.
- Supporting Maintenance and Reliability as an integral part of business and asset management.
- Presenting a collective voice on Maintenance and Reliability issues and advancing innovative reliability practices.
- Facilitating information exchange through a structured network of Maintenance and Reliability professionals.

2 MEMBER BENEFITS

-Educational Workshops

-Benchmarking and Standards

-Quarterly Newsletter

-Local Chapters

-Networking Database

-Advertising Opportunities

-Web Site and Discussion Boards

-Annual Conference

-Executive Member Meetings

-Membership Directory

-Speaker's Bureau

3 TYPES OF MEMBERSHIPS

Individual: at $125/person. Classifications include: practitioners, suppliers, consultants, students, military, academic, government, and non-profit trade organizations.

Executive: at $1,250/year. Include full membership privileges to five designated individuals. Executive memberships include free participation in Annual Benchmarking program, one vote at Annual Business Meeting, access to job posting program and other key benefits.

4 GOVERNANCE

The governing body is as depicted below:

2005-06 Board of Directors

Officers		**Directors**	
Chair	Chuck Armbruster	Define Body of Knowledge	Ramesh Gulati
Vice Chair	Tom Byerley	Education	Chad Sayle
Treasurer	Tim Goshert	Certify & Set of Standards	Dane Brooks
Secretary	David Staat	Improvce Members Services	John Schultz
Past Chair	Larry Cote		

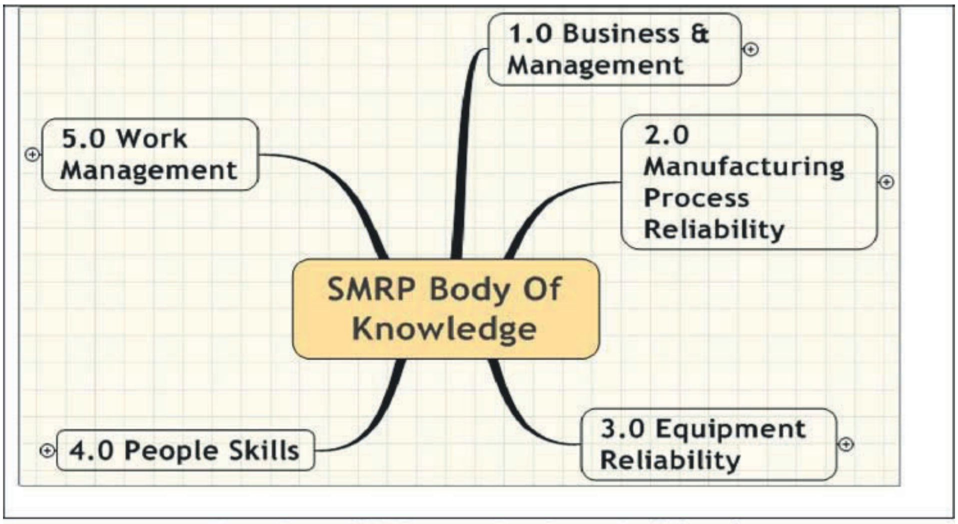
5 SMRP INITIATIVES

SMRP's leadership is comprised primarily of Maintenance & Reliability practitioners whose dedication and commitment to facilitating networking and the exchange of knowledge is impressive and unquestioned. Individually and collectively, the leadership is responsible for ensuring that SMRP's strategic initiatives are tied to our mission, purpose and vision, and that they are being implemented with the utmost sensitivity to quality, usefulness, timeliness and relevance.

These current initiatives include these four areas:

5.1 Define the Body of Knowledge

- Benchmarking- Seeking out the best practices that can be adopted to improve performance
- Best Practices- Developing Maintenance and Reliability metrics utilizing common terminology
- Body of Knowledge- Defining the Maintenance and Reliability professional. The subject areas include the areas found at the top of page:



Body of Knowledge Initiative

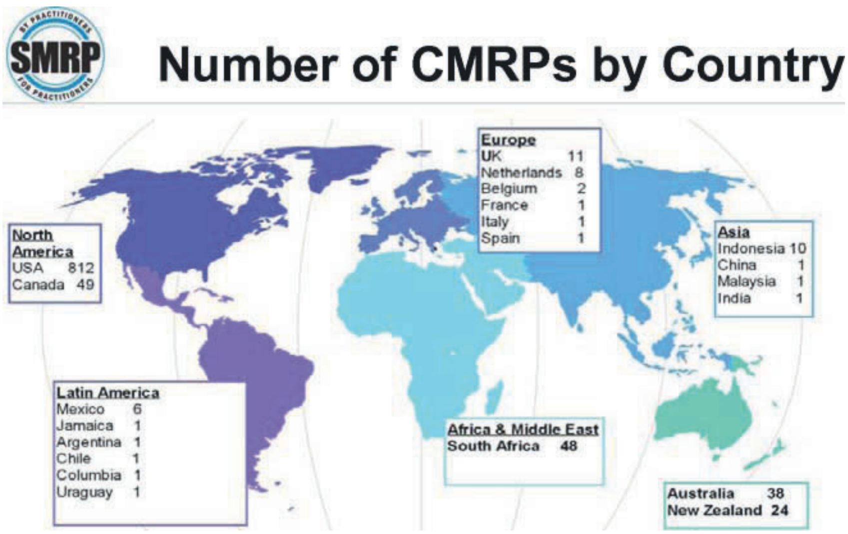
5.2 Educate

- Providing workshops and Executive Member Meetings
- Acting as a liaison between the private sector and educational institutions
- Executing outstanding annual conferences
- Developing the SMRP web site as the go-to resource for Maintenance and Reliability content

The SMRP Home Page (www.smrp.org) is shown below:

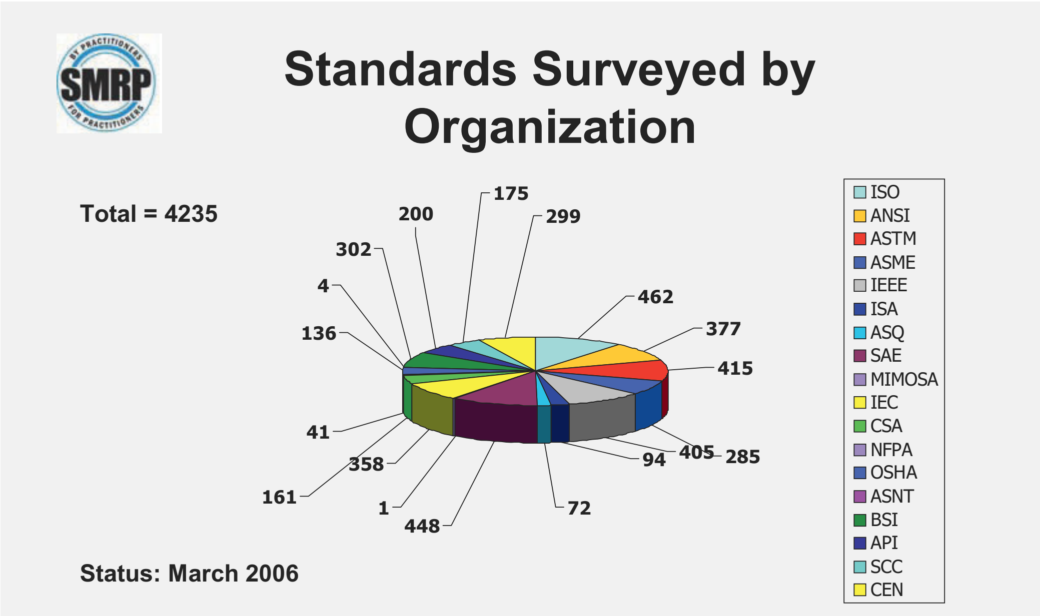
5.3 Certify & Set Standards

- SMRP Certifying Organization (SMRPCO)
  ** Prepares and conducts examinations for Certified Maintenance & Reliability Professional (CMRP) recognition
  ** The SMRPCO exam covers the five subject areas of BoK
  ** The Exam questions are developed by practitioners and statistically analyzed



5.4 Standards Committee

** Reviews, catalogs, and develops standards that affect the Maintenance and Reliability community
** Applying to ANSI for accreditation as a Standards Development Organization



5.5 Improve Member Services

- Enhancing membership value
- Promoting membership growth
- Establishing local chapters

MODELING RAIL WHEEL-FLAT DYNAMICS

NS Vyas & AK Gupta

Indian Institute of Technology, Kanpur 208 016, India

Abstract: Development of flats on the wheels of railway engines and wagons is one of the major causes of track and ride quality deterioration. Wheel flats are flat zones on the wheel tread caused by unintentional sliding of the wheel on the rail when the brakes lock. A fresh wheel-flat has the shape of a geometric chord on the wheel circumference. Forces from the wheels with flats can be up to four times higher than those from normal wheels. In this paper, wheel-flat and rail interaction dynamics is modeled through a finite element model. Typical wheel flat loading patterns are generated and response is computed for various cases of rigid and flexible rail supports and damping.

Key Words: Dynamics, Rail Wheel, Finite Element Model, Damping.

1 INTRODUCTION

A wheel flat is caused by the brake or some other mechanism, interfering with the normal angular velocity of the wheel. When brakes do not function properly, the wheel ceases to rotate and a flat develops due to plastic deformation. The dimensions of the wheel flat, its form and position on the tread depends on length of the wheel skid, axle loading, running speed, shape of rail profile, position of the wheel on rail, wheel material and the material coefficient of adhesion between the wheel and the rail material.

Wheel flats cause large dynamic forces on the rail. According to popular practices, flats on wheels should be restricted to a length of 60 mm and a depth of 0.9-1.4 mm. In this range the dynamic wheel load increases by approximately, 30 kN/mm for timber sleepers and 50 kN/mm for concrete sleepers. Wheel flats are also known to inflict very high acceleration levels (about 500g) on the track. These high vibrations get transmitted to the rolling stock, which induce forces much larger than permissible values and cause damage to the suspension system, frame and the body of the rolling stock.

The general problem of rail-wheel interaction attracted attention as early as the year 1847, when a British commission was appointed to study dynamic effects on railway bridges. The differential equation of deflection was established by neglecting inertia of the bridge and by considering the moving load as a concentrated mass. An exact solution of the equation was obtained by Stokes [1]. Timeshenko [2] solved the problem of a pulsating forces moving with uniform velocity along a beam. Lowan [3] studied the case with variable velocity of traversing force. These investigations pertained to single span structures. Ayre, Ford, and Jacobsen [4] studied transverse vibration of a two - span beam under action of a moving constant force. Some of the more recent works include a study carried out by Cai, Cheung and Chan [5] who analysed the dynamic response of an infinite continuous beam subjected to a moving force. For multi-span beams most researchers have used Finite Element methods. Nielsen and Igeland [6] have investigated the vertical dynamic behaviour for a railway bogie moving on a rail by use of an extended state-space vector approach in conjunction with a complex modal superposition for the track. Effects of imperfections on the running surface of wheel and rail were studied by assigning irregularity functions to these surfaces and contact forces were calculated for moderate vehicle speed in the presence of wheel flat. Thambiratnam and Zhuge [7] developed a simple finite element model to analyze simply supported beams on an elastic foundation of any length, where the foundation has been modeled by springs of variable stiffness. Abu Hilal and Zibdeh [8] investigated vibration of elastic homogeneous isotropic beam with general boundary condition traversed by moving force. Wu and Shih [9] have studied dynamic response of railway and carriage under the high speed moving loads. Dynamic response of the rail and carriage due to action of multi-roller carriage was determined by means of the finite element method. Wu, and Thompson [10] have investigated vibration analysis of railway track with multiple wheels on the rail.

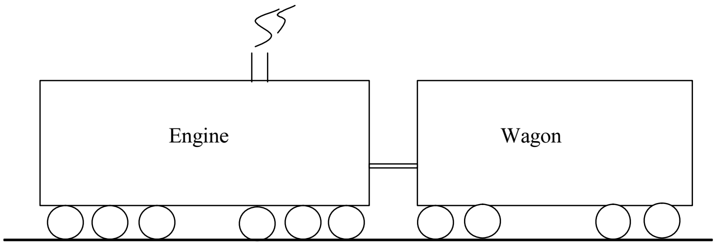
In this paper wheel-flat and rail interaction dynamics, modeled through a finite element model is described. Typical wheel flat loading patterns are generated and response is computed for various cases of rigid and flexible rail supports and damping.

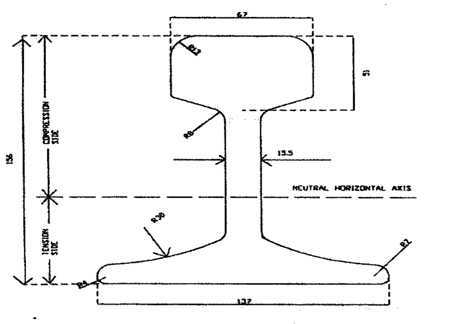
2 MODEL AND ANALYSIS

Rail vibration under loading due to normal wheels is estimated initially and then modeling is extended to incorporate loading due to wheels, which have developed flats. Modeling is done through the Finite Element package NISA and the wheel loads are treated as moving loads.

2.1 Rail Data

Analysis is carried out for typical data obtained from Indian Railways. A railway train comprising of an engine and one wagon is considered. A schematic of the train is given in Fig.1 (a). The engine has 12 wheels (six wheels on either side). The wagon has eight wheels, four on either side. Analysis is carried out for only one side of the railway track. The rail cross-section is shown in Fig. 1 (b). The data employed for the analysis is listed below:





(All dimensions are in mm)

Fig 1 (a) Schematic diagram of test train

Fig 1 (b) Rail section

Cross section area of rail	66.15 x 10-4 m^2
Moment of inertia	2158 x 10-8 m^4
Distance between two-sleeper pads	0.6 m
Young modulus of rail	2.0678 x 1011 N/m^2
Density of rail	7600 Kg/m^3
Speed of train	36 km/h (=10m/s)
Wheel diameter	1 m
Load on each wheel of engine	92214 N
Load on each wheel of wagon (empty)	29062.125 N
Total Number of wheels in the engine	12
Total Number of wheels in the wagon	8
Damping ratio of rail (rigid support)	0.05
Stiffness of sleeper pads	2.5 x 108 N/m
Damping ratio of rail (flexible support)	0.1

The weight of the engine and wagon is taken to act on the rail through the wheels as point forces. The respective weights of the engine and wagon are assumed to be equally distributed among the wheels. At the initial time instant the first wheel of the engine is assumed to be at node 1 of span number 1 of the rail. Table 1 gives the wheel loading data for all the ten wheels (6 from engine and 4 from the wagon).

Table 1

Wheel loads, distances and arrival times at node 1 (Train velocity =10 m/s)

Wheel No.	Arrival time at node 1 (s)	Wheel load (kN)	Distance from first engine wheel (m)
1	0.00	92.214	0.0
2	0.21	92.214	2.1
3	0.42	92.214	4.2
4	1.05	92.214	10.5
5	1.26	92.214	12.6
6	1.47	92.214	14.7
7	1.79	29.062	17.9
8	1.99	29.062	19.9
9	2.44	29.062	24.4
10	2.64	29.062	26.4

Total time taken to cross rail by train = (rail length + train length) /velocity = (9 + 26.4)/10 = 3.54 s

2.2 Finite element model

Two models of rails are developed - (i) rail with rigid supports (ii) rail with flexible supports. A rail track of 15 spans is considered. The length of each span is 0.6 m. Different scenarios with wheel flat falling on the supports and in-between supports, are discussed. Details of the FE model are given below:

(i) Rail with rigid support

Type of element	2-D beam
Order of element:	1 (linear)
Degrees of freedom per node	3 (u_x,u_y,r_z)
Number of nodes per element:	2
Total number of elements in rail	118
Total number of nodes in rail	119
Total length of rail	9 m

u_x = 0 is taken at all nodes of rail and u_y = 0 at supports only

(ii) Rail with flexible supports (in addition to the data of case (i))

Type of element of spring	2-D translation
Order of spring element	1(linear)
Degree of freedom spring element	2 (u_x,u_y)
Number of nodes per spring element	2
Total number of spring elements	15

u_x =0 is taken at all nodes of rail and u_y =0 at one end of spring element only. The above models are shown in Fig 2 (a) and Fig 2 (b) respectively.

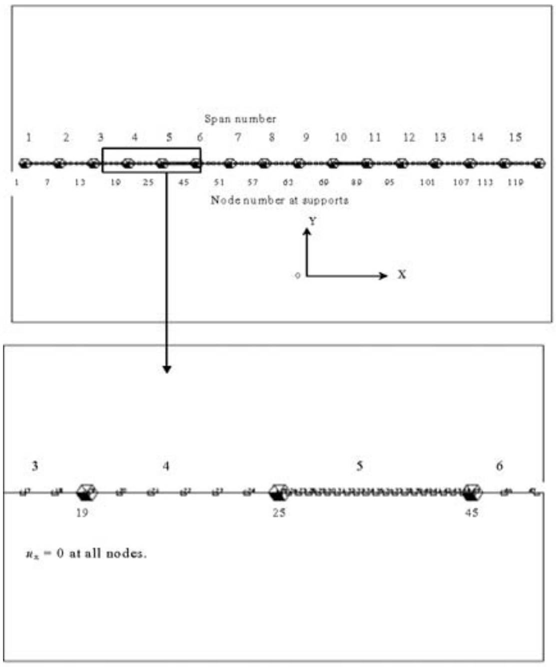


Fig 2(a) Finite Element model of 15-span rail-rigid supports

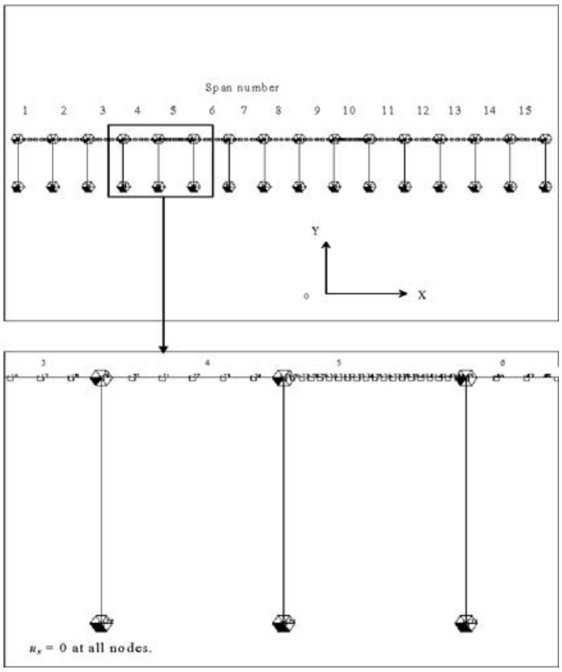


Fig 2(b) Finite Element model of 15-span rail-flexible supports

2.3 Free vibration characteristics of rail

Eigenvalue analysis of the rail was carried out in NISA using conventional subspace iteration and consistent mass formulation. Table 2 gives the first five natural frequencies of rail for the rigid support and flexible support models. The corresponding mode shapes are given in Figs. 3 (a) and (b).

Table 2

First five natural frequencies of rail (rad/s)

Mode No.	Rigid support	Flexible support
1	7842.73	2854.00
2	7941.44	2854.11
3	8230.34	2862.73
4	8690.14	2891.54
5	9295.03	2955.22

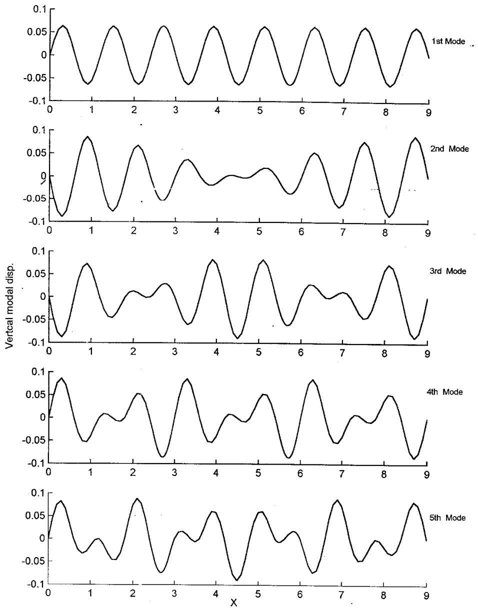


Fig 3(a) Mode shapes of rail – rigid supports

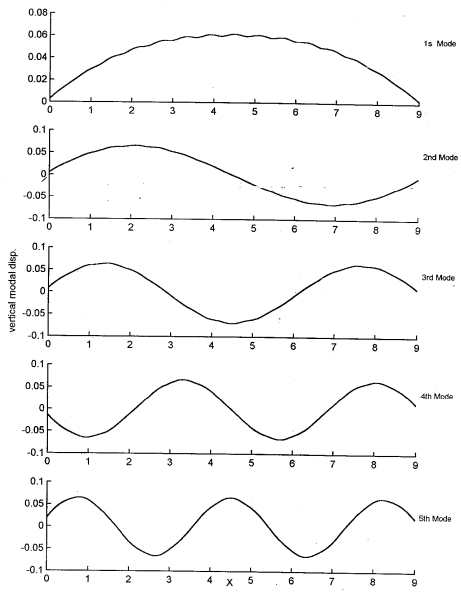


Fig 3(b) Mode shapes of rail – flexible supports

2.4 Wheel flat loading

The loading pattern caused due to out-of-roundness of the wheel depends on the severity and shape of the damage on the wheel. The out-of-roundness is however idealised by a flat on the surface, which would be seen as a chord on circular profile of the wheel. Such a flat results in loss of contact between the wheel and the rail for a short time and then regaining of contact. This idealised lift and hit phenomenon is shown in Fig. 4. When there is no contact, the wheel rises above the rail. Recovery of contact results in a large impact force. However, due to the bogie suspension system, the loss of contact between the wheel and the rail is not total, since the wheel moves downwards and the rail moves upwards. The overall effect is that the contact surface initially decreases and later, due to inertia, the wheel continues downward motion resulting in an increase of contact force.

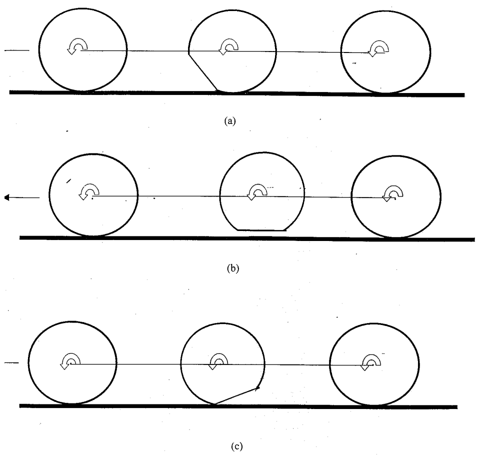


Fig 4 Lift and hit phenomenon of wheels

2.5 Rail vibration under moving train

Deflection and acceleration are computed for point loads from the ten moving wheels for the following cases:

(i) Undamped System (a) without flat (b) with flat
(ii) Damped System (a) without flat (b) with flat.

The wheel flat is assumed to be present on the 10th wheel of the train. Chord length of the flat is taken as 60 mm. Forced vibration analysis is carried out in NISA and 30 modes are taken under consideration. The wheel-flat hits the rail on the 5th and then on the 10th span. In order to improve accuracy 20 elements are taken on these spans, while on the remaining spans the number of finite beam elements are 6. Thus, the response is computed for 119 nodes on the rail. The loading patterns for cases with and without wheel flats are shown in Fig.5 for 10th span. The flat wheel looses contact at node 33 and regains it at node 35. Again, it looses contact at node 82 to regain it at node 84.. The nodal distances and arrival time of the 10th wheel at these nodes along with the tenth node (which sees normal wheel behaviour) are listed in Table 3.

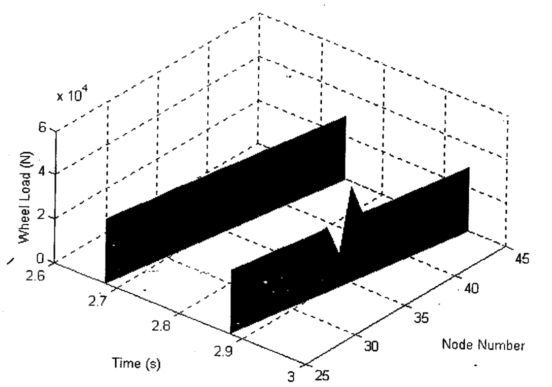


Fig 5 Load pattern of 10th span under 9th, 10th wheels

Table 3

Nodal distances and arrival time of 10th wheel

Node Number	Distance from Node 1 (m)	Arrival time of the 10th wheel (s)
10	0.90	2.730
33	2.64	2.904
34	2.67	2.907
35	2.70	2.910
82	5.79	3.219
83	5.82	3.222
84	5.85	3.225

Node 35, where the hit occurs for the first time, lies in the centre of a beam span (number 5). Node 84, where the hit takes place for the second time is close to a support (i.e. a sleeper). The vibration of rail is expected to be more due to the first hit, in comparison to the second hit. For the undamped system with rigid supports the displacement response at nodes 10, 33 and 35 is shown in Figs. 6 (a)-(c). It corresponds to the normal case (without flat). Figs 7 (a)-(c) give the response in the case of a wheel with flat. The corresponding acceleration responses are shown in Figs. 8 (a)-(c) and 9 (a)-(c). It can be seen that while each wheel causes distinct deflection as it passes through each node, the acceleration response does not provide a clear picture since it comprises of a large number of frequency components. It can be said that measurement of displacement response through strain gauging techniques can provide better information than that obtained through measuring acceleration, using accelerometers.

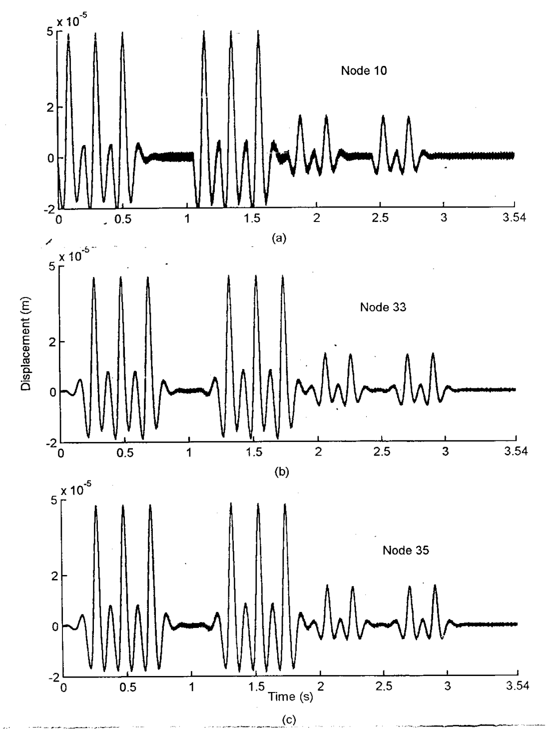


Fig 6 Rail response for Case 1(a) Rigid supports (a) node 10, (b) node 33 and (c) node 35

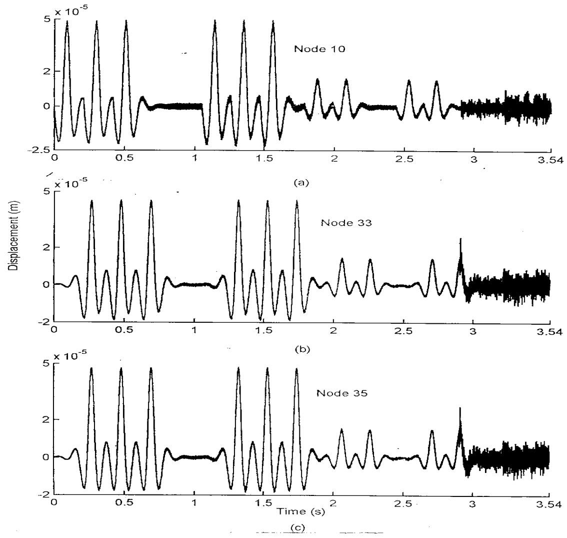


Fig 7 Rail response for Case 1(b) nRigid supports (a)node 10, (b) node 33 and (c) node 35

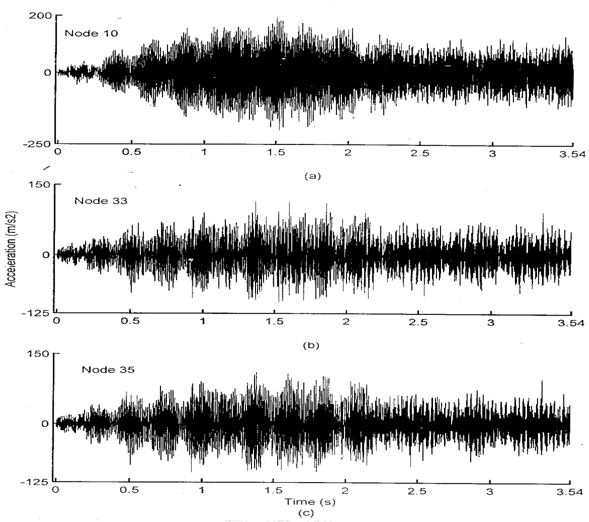


Fig 8 Acceleration response for Case 1(a)Rigid supports (a) node 10, (b) node 33 and (c) node 35

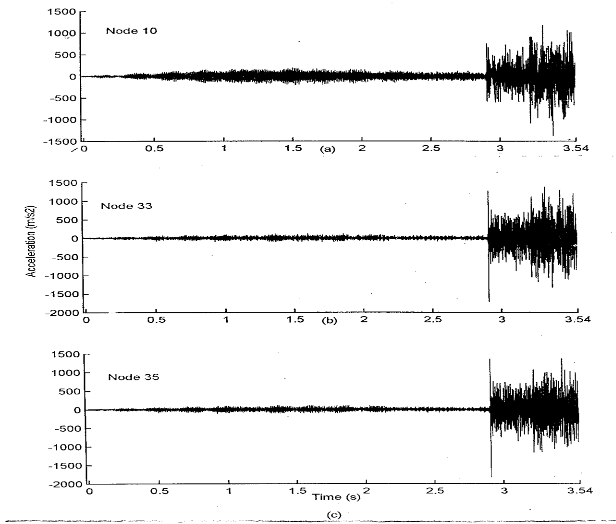
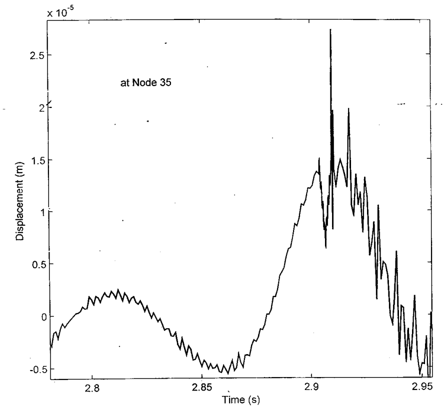


Fig 9 Acceleration response for Case 1(b) Rigid supports (a) node 10, (b) node 33 and (c) node 35

From the plots of Figs. 7 (b), the impact caused by the wheel flat is clearly visible. The maximum deflection due to the engine wheels is 47.82 μm at node 35. Maximum deflection, at node 35, due to the wagon wheels is 15.59 μm in the absence of any flat, while it rises to 27.35 μm, when a flat is present. Occurrence of this increase in the maximum deflection can serve as an effective tool for wheel-flat detection. The zoomed picture of the response under wheel flat loading is depicted in Fig.10, which clearly indicates the lift and hit mechanism of the wheel flat discussed earlier.



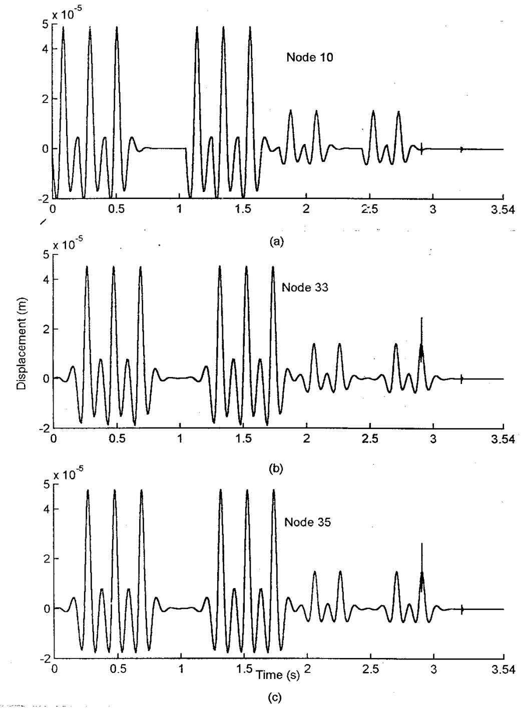
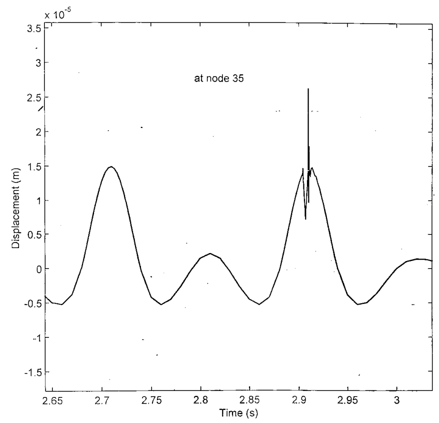


Fig 10 Zoomed view of response for Case 1(b) Rigid supports

Fig 11 Rail response for Case 2(b) Rigid supports (a) node 10, (b) node 33 and (c) node 35

The response for the damped system is shown in Figs. 11 (a)-(c). The inclusion of damping in the analysis, as can be seen, removes the free vibration components, from the response plots. The zoomed picture of the response under wheel flat loading is given in Fig.12. Computation is also carried out for cases where the rail support has been taken as flexible. The response in this case is shown in Figs. 13 (a)-(c). As expected, deflection is higher in this case than with rigid supports. The response at nodes 82, 83, 84 (for rigid supports and with damping) is shown in Figs. 14 (a)-(c). Comparison with the response at nodes 33, 34, 35, discussed earlier reveals that the deflection levels are lower and the impact due to wheel flat is also less pronounced. This is due to the fact that while nodes 33,34,35 lie in the centre of a beam span, nodes 82, 83, 84 are closer to a support (i.e. sleeper).



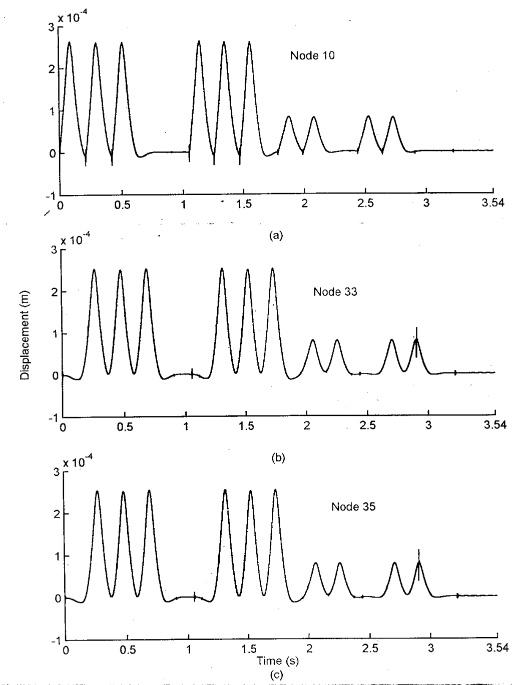


Fig 12 Zoomed view of response for Case 2(b) Rigid supports

Fig 13 Rail response for Case 2(b) Flexible supports (a)node 10, (b) node 33 and (c) node 35

The maximum deflections caused by the wagon wheel, in various cases discussed above are listed in Tables 4.

Table 4

Maximum deflection due to wagon wheel (at node 35) in μ m

Type of support	Without flat		With flat	
	Undamped	Damped	Undamped	Damped
Rigid support	15.59	14.89	27.35	26.26
Flexible support	221.26	79.67	203.00	109.75

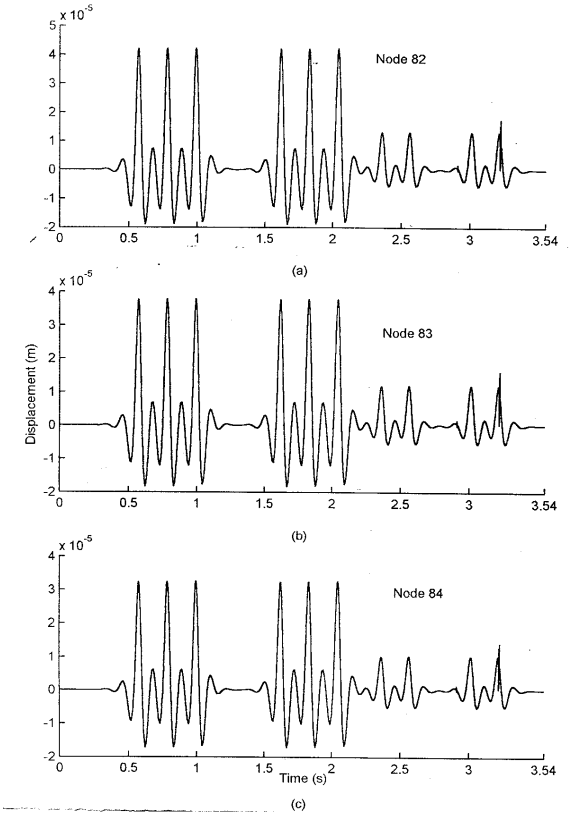


Fig 14 Rail response for Case 2(b) Flexible supports (a) node 82, (b) node 83 and (c) node 8

3 CONCLUSION

The analysis reveals that flat zones on the rail wheel periphery cause impacts on the rail with a magnitude nearly twice as much as the normal rolling force. It is important to identify these deformities and isolate such wheels quickly in order to prevent rail fractures.

4 REFERENCES

[1] Stokes, Sir G. G.(1849), "Discussion of a Differential Equation Relating to Breaking of Railway Bridges", *Mathematics Physics Paper*, Vol. 2, 179,.

[2] Timoshenko, S. P., "On the Forced Vibration of Bridges", *Philosophical Magazine*, Vol. 43, p. 1018, 1922.

[3] Lowan, A. N.(1935), "On Transverse Oscillations of Beams Under the Action of Moving Variable Loads", *Philosophical Magazine*, Vol. 19,series 7, 708.

[4] Ayre, R.S., Ford George and Jacobsen, L.S.(1950), "Transverse Vibration of a Two Span Beam Under Action of a Moving Constant Force", *Journal of Applied Mechanics*, Vol. 17, 1 - 12.

[5] Cai, C. W., Cheung, Y .K., and Chan, H. C.(1988), "Dynamic Response of Infinite Continuous Beams Subjected to a Moving Force", *Journal of Sound and Vibration*, Vol. 123, 461 - 472.

[6] Nielsen, J. C. O. and Igelend, A.(1995), "Vertical Dynamic Interaction Between Train and Track - Influence of Wheel and Track Imperfections," *Journal of Sound and Vibration*, Vol. 187, 825 – 839.

[7] Thambiratnam, D. and Zhuge, Y.(1996), "Dynamic Analysis of Beams on an Elastic Foundation subjected to Moving Loads," *Journal of sound and vibration,* Vol. 198, 149 – 169.

[8] Abu Hilal, M. and Zibdeh, H. S.(2000), "Vibration Analysis of Beams with General Boundary Conditions Traversed by a Moving Force," *Journal of Sound and Vibration*, Vol. 229, 377 – 388.

[9] Wu Jong-shyong and Shih Po-Yun (2000), "Dynamic Response of Railway and Carriage Under the High-speed Moving loads," *Journal of Sound and Vibration*, Vol. 236, 61 – 87.

[10] Wu, T. X. and Thompson, D.J.(2001), "Vibration Analysis of Railway Track with Multiple wheels on the Rail,' Journal of Sound and Vibration, Vol. 239, 69 – 97.

TOWARDS AN ASSESSMENT TOOL FOR THE STRATEGIC MANAGEMENT OF ASSET CRITICALITY

D Anderson [a] P Keleher [a] and Peter Smith [a]

[a] *Faculty of Sciences, Engineering and Health, Central Queensland University, Queensland, Australia*

Abstract: Many organisations in both the private and public sectors rely on engineering assets to provide products and services. The failure of these assets, or their components, can have undesirable consequences for the organization. Failure consequences can be measured in terms of their impact on the safety of personnel, the public and the environment, and on their impact on desired levels of service or output. Existing decision tools for determining the criticality of asset failure are typically rigorous and detailed. They are necessarily time consuming methodologies. A strategic management approach seeks to focus the effort and cost of such detailed analysis more appropriately. It recognises the value of the expert knowledge of those directly engaged in operations and maintenance functions in supporting a strategic approach to the assessment of asset criticality. The purpose of such a tool is to quickly and effectively identify critical assets within a facility to assist strategic decision making. The assessment tool being trialled is to support the Maintenance function of organisations and therefore it is focussed on the Operations and Maintenance phase of the asset life cycle. This paper presents an overview of the alternative methodology being trialled and some examples of its practical application.

Key Words: Asset, Failure, Consequence, Criticality, Probability, Risk, Methodology

1 INTRODUCTION

Expert knowledge based assessments of asset criticality are necessarily qualitative. They are best supported by a systematic and standardized methodology that underpins qualitative rigor and enables assessment to be open to external audit. This paper (a) explores the concept of asset criticality, (b) describes the systematic approach to qualitative assessment that is being piloted, (c) outlines the results of the pilot study series completed and (d) draws some conclusions and makes some suggestions about future steps in the further development of expert knowledge based assessment tools of asset criticality.

2 THE CONCEPT OF ASSET CRITICALITY

Asset criticality can be determined asset by asset as part of a complete quantitative diagnosis of current state of the asset, its potential for future failure and the consequences of such failure. It may also be assessed in order to choose or sort assets into a priority list. The consideration of asset criticality in this paper is in the context of this latter, triage like, approach with a focus on prioritizing maintenance work to support asset operation and safety.

Asset intensive organisations rely on physical assets to produce products or services for end users or customers. The physical assets used by these organisations include industrial assets (plant and equipment) and infrastructure assets (drainage, water and sewer systems, bridges, roads, tunnels, and dams).

There are similar objectives identified for asset intensive organisations typified in statements by Rio Tinto [1] and Shell [2]:

- To operate profitably or economically
- To provide desired or agreed levels of service
- To operate safely in terms of consumers, the public and employees
- To operate safely with regard to the environment.

Asset performance in asset intensive organisations directly impacts on the ability of the organisation to meet objectives. The loss of asset capability through degradation or failure can impact significantly on the degree to which an organisation is able to meet their objectives. The Maintenance Engineering Society of Australia's Capability Assurance Model [3] draws a direct relationship between asset capability (function) and the achievement of organisational objectives. Kelly [4] describes the loss of asset function affecting the production and safety performance of a facility.

Asset criticality is defined in literature in different forms. Moteff and Parfomak [5] draw on numerous US sources to identify critical assets as those assets whose 'incapacity or destruction would pose a threat to security, national economic security, or public health or safety'. This definition is a more general view of asset criticality and less focussed on the physical failure of assets, rather a loss of asset function by any means.

An operations and maintenance view of critical assets is defined in the International Infrastructure Management Manual [6] as:

Assets for which the financial, business or service level consequences of failure are sufficiently severe to justify proactive inspection and rehabilitation.

This definition of critical assets is focussed specifically on decisions surrounding preventive maintenance actions arising from perceived modes of asset failure.

There are potentially many possible viewpoints of asset criticality. Examples of these viewpoints derive from the lifecycle phase of the asset [7] and organisational function that perceives asset criticality [8]. Models that describe asset criticality from all possible perspectives could be complex. The scope of the assessment in this paper is concerned with a perspective of the Operations and Maintenance phase of the asset life cycle and a view from the perspective of the Maintenance function of the organisation.

For the purpose of the initial study, the following definition of asset criticality is proposed:

A relative measure of the impact of the loss of asset function on the objectives of the organisation from the viewpoint of operations and maintenance.

This definition allows for asset criticality to be modified according to local business rules. A range of vulnerabilities is possible within the scope of the definition. This is considered to be a working definition. It is expected that this definition will be refined with experience of assessing criticality and feedback from industrial facilities.

3 CURRENT METHODS FOR IDENTIFYING CRITICAL ASSETS

There are few published methodologies for identifying critical assets. A search of current literature has identified frequent references to the need to identify critical assets, but little on a methodology for identifying critical assets [4, 5, 9, 10, 12].

International Standard IEC 60300-3-14 [9] advocates a preliminary 'Operation Use Study' as a precursor to developing maintenance policy. This study is to determine:

- items and systems that are important to meet process objectives
- items and systems that involve safety considerations
- operational patterns
- environmental conditions
- expected service life

There is no methodology or reference within the standard that describes how such a study might be performed or documented.

Kelly [10] describes a 'Top-Down' approach to identifying critical assets as a precursor to developing maintenance policy. The following steps to ranking critical assets are identified:

a) Construct a process flow diagram and establish a plant inventory
b) Understand the plant operating characteristics and the production policy
c) Rank units is order of their importance (criticality)

Kelly does not provide a methodology or reference that describes how such a study might be performed or documented.

There is evidence of asset criticality being based purely on failure consequence as used by Shell Brunei [11] to rank asset criticality. It is more common for a risk based analysis based on an assessment of asset failure consequence and probability [12] to be applied. Where failure probability is considered at component and failure mode level, as typified by Failure Modes and Affects Analysis (FMEA) [13], estimates of probability are valid. It is otherwise considered difficult to assess how probable a loss of asset function event is without a rigorous analysis of the way in which such a loss of function might occur.

Where an underlying principle is provided in support of the consequences of the loss of asset function, it is usual that criteria based assessment is applied. These criteria vary between approaches. A review of methods [8, 10, 9, and 11] identified the following criteria for assessing the consequences of loss of asset function:

- Repair costs
- Loss of income

- Loss of service
- Loss of life or injury
- Damage to property
- Failure to meet statutory requirements
- Third party losses
- Loss of image
- Location of asset
- Environmental Impact

There are two distinct categories of criteria identified from this list. These are 'mandatory' criteria dealing with Safety, Environmental and Statutory requirements and 'economic' criteria dealing with loss of income, service or damage to property.

The method of ranking asset criticality within the criteria is also varied across the methods reviewed. This is summarised by Jay [8], who identifies three common approaches to ranking asset criticality:

- High, medium, low;
- Scoring (e.g. 1,2,3,4,5);
- Weighted scoring.

Jay regards weighted scoring as being a more discriminatory and favoured method.

There is not one single approach in the methods reviewed that is considered to meet the requirements of defining critical assets as defined in this paper. There is an opportunity to draw on the strengths of current methods to create an alternate methodology for identifying critical assets.

4 AN ALTERNATE METHOD FOR IDENTIFYING CRITICAL ASSETS

4.1 General Description

The aim of developing an alternate method for identifying critical assets was to draw on individual features of existing methods to create an approach to:

- Identify assets;
- Rank critical assets;
- Be easily applied and reproduced;
- Take an operations and maintenance perspective of critical assets.

The following elements of a methodology were developed:

Asset Identification – a means of identifying and representing assets for the purpose of criticality analysis.

Criteria – a list of factors against which asset criticality could be measured.

Weighted Scoring – a means of allocating scores to achieve a total score indicating asset criticality.

Scoring Guides – templates for the application of scoring against each of the identified criteria.

Application – a means of applying the scoring to the criteria within an organization.

These elements are explained in more detail in the following sections.

4.2 Asset Identification

Assets were identified visually and models prepared. Assets were clustered in groups according to logical units or process function. This satisfied the layout of the facility, and also assisted with segregating assets for analysis purposes, described in more detail in section 4.5. Asset identification is demonstrated in the block model of Figure 1. In this example assets are represented as blocks interconnected with lines representing logical process flows or operational or reliability relationships.

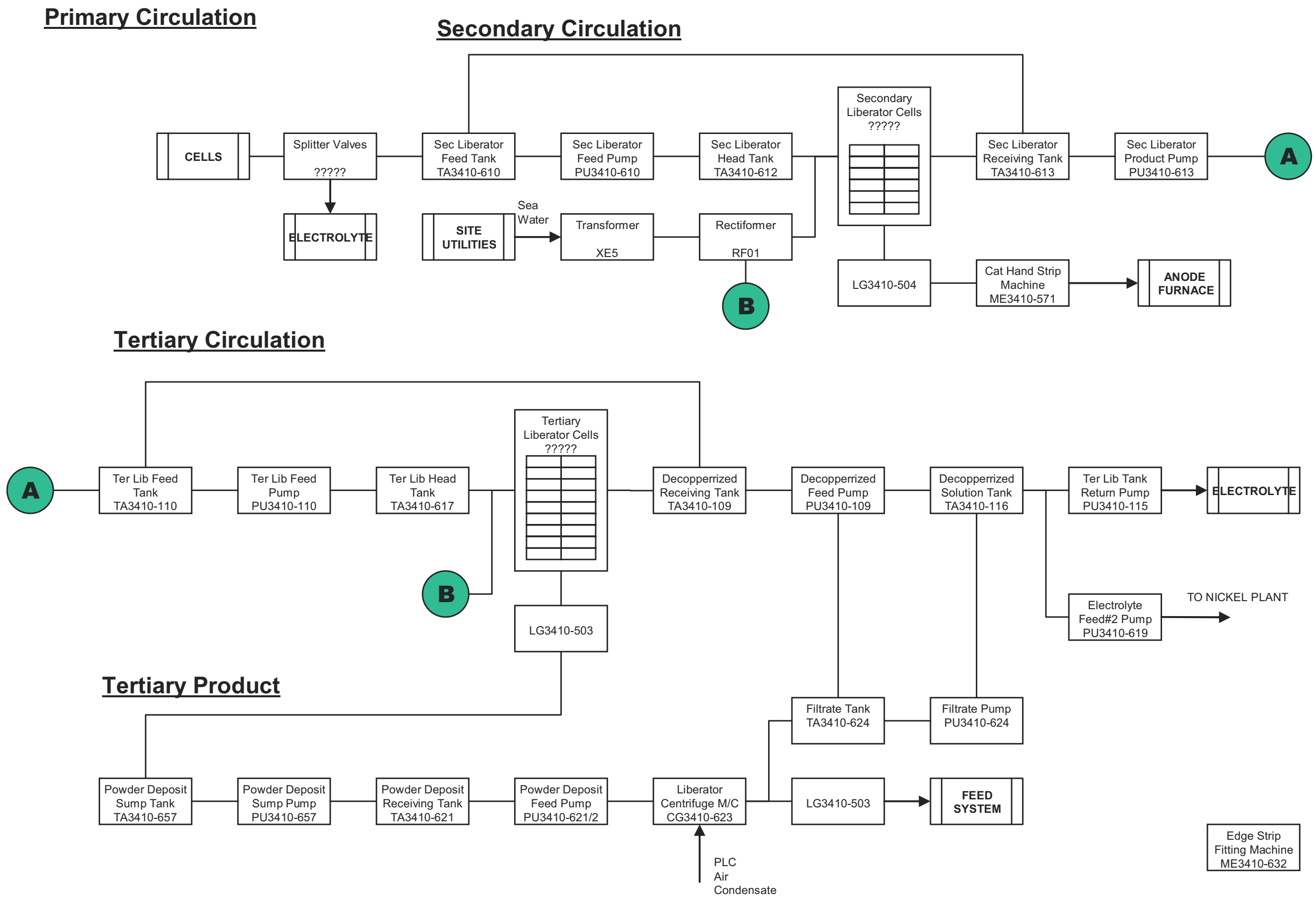


Figure 1- Asset Representation

Initial modelling of assets was performed based on recognising logical groups of components with a clear input and output or a definable function. It was also considered important to clearly identify stockpile or storage areas, inputs and outputs to the assets and links to other systems within the scope of the assessment.

4.3 Criteria

A criteria assessment using a weighted scoring method was selected for trial based on current methods. Initial criteria were based on the key business drivers of Health, Safety and Environment (HSE) and Output were readily identified from the function of maintenance. There was little logical basis for a ranking on two criteria. Additional criteria to assist in ranking based on probability were added to the criteria set. The criteria initially selected were:

Safety / Environment (mandatory) - the consequences of loss of asset function on the safety of personnel, the public, consumers or the environment.

Output (economic) - the consequences of loss of asset function on the loss of service levels, or the loss of revenue through termination of production, lost opportunity, loss of product or cost of reproduction or reparation.

Severity - the potential rate that the environment, process conditions or constituents can deteriorate assets.

Condition - the actual or assumed condition of the plant as determined by actual historical performance where possible.

Existing Maintenance - the actual or assumed quality of preventive maintenance actions carried out on the asset.

Complexity - the complexity of the asset according to the number of maintenance causing items that make up the asset.

A system of weighted numerical scoring was developed to indicate the degree of impact that each criteria had on the perceived criticality of each asset.

4.4 Weighted Scoring

An arbitrary scale of 1 to 10 for scoring and weighting was selected. It was assumed that a factor of more than 10 would be unnecessary to discriminate between critical assets, and that any less than 10 would be insufficient to achieve the desired spread of results.

Criteria weighting was varied to suit local business rules. This depended on the emphasis of the organisation (e.g. relative safety vs. process criticality) and the reliability of information used to conduct the assessment so that the results are not greatly influenced by information from unreliable or biased sources. Weighting was maintained separately to scoring so that results could be compared across facilities or sensitivity analysis performed.

Low scores indicate a less serious outcome within the criterion. A score of 1 indicated a low criticality against a criterion while a score of 10 indicated the highest. A weighting of 1 indicated the lowest importance of a criterion to the ranking, while a score of 10 represented the highest. Scores and weightings in between 1 and 10 indicate relative scales for comparison purposes.

The final ranking of the criticality of assets is determined by the total weighted score. The sum of each score multiplied by the criterion weighting for all criteria provided the total weighted score for a single asset. The magnitude of the total weighted score for the asset determined its overall ranking within the facility.

4.5 Scoring Guides

Scoring guides were developed to apply consistent application of scoring to the criteria. Although not covering every possible description of a criterion outcome the guides are intended to maintain a consistent scoring across a facility with regard to exposure and consequences.

Table 1- Scoring Guide - Safety / Environmental Criterion

Score	Description
1	No impact on environment or safety of personnel if failure occurs.
2	Nuisance impact on employees and no impact on environment if failure occurs e.g. noise, additional rectification required.
3	Nuisance impact on employees and impact on environment recoverable at minimum cost.
4	Failure results in an obvious hazardous situation but contained by equipment. e.g. Oil spill within bunded area.
5	Failure results in obvious hazardous situation, but not contained by equipment. Hazard can be seen and avoided or approached with caution. e.g. Dust leak.
6	Failure results in hidden unsafe equipment or working environment (Hazard is not visible).
7	Serious hazard to individual employee or release of environmentally sensitive substance.
8	Potential fatality of individual employee or release of environmentally sensitive substance. e.g. Lifting device failure.
9	Serious hazard to multiple personnel or wide spread environmental impact. Hazard cannot be contained or controlled. e.g. Fire.
10	Potential for fatality of multiple personnel or widespread environmental consequences. e.g. Major Chlorine Gas Leak.

Table 2- Scoring Guide - Output Criterion

Score	Description
1	No affect on production within repair time
2	Reduction in production capacity short time frame (without affecting ability to meet orders).
3	Loss of equipment with some contingency and buffer storage
4	Loss of equipment with significant contingency and buffer storage.
5	Loss of equipment with 100% capacity backup.
6	Failure of equipment without sufficient back up to ensure that production requirements can be met. 25% loss of facility output or equivalent loss.
7	Significant impact on ability to meet customer order quantities. 50% loss of facility output or equivalent loss.
8	Significant impact on ability to meet customer orders for specific product with no backup. 75% loss of facility output or equivalent loss.
9	Significant impact on multiple equipment or process lines. Equivalent 75% loss of output of facility.
10	Stops entire output of facility. Contributes to long outages of facility. e.g. Loss of power.

Similar scoring guides were developed for the remaining four criteria. These scoring guides are omitted from this paper due to size consideration.

4.6 Application

The preliminary steps in the application of the criticality assets were performed by a facilitator with input of management stakeholders and the assistance of a facility representative. This included:

1. Identifying and documenting asset models;
2. Building an asset list;
3. Establishing the criteria;
4. Establishing and ratifying scoring guides;
5. Weighting criteria.

Once the preliminaries had been performed, the scoring was applied in a facilitated meeting by knowledge workers from the facility under analysis. The mix of individuals in the team was sometimes varied during the application; particularly on larger sites where broad expertise was not available within individuals.

Scores were applied to the criteria for each asset. A pessimistic application of the scoring guides was applied, based on the knowledge and understanding of the participants. Discussion and consensus were sought before each asset was considered to have been assessed. The outcome of the scoring was captured as indicated in Table 3. Initial analyses were captured manually, but this was soon replaced by a computer aided data collection process.

Table 3 - Scored Asset

UNIT - COMPRESSOR
ASSET - SEAL OIL SYSTEM

	Weight	*Score*	*TWS **	*Comments*
Safety / Environment	10	10	100	Potential for multiple fatalities
Output	8	10	80	Stops Entire Output of Facility
Severity	7	2	14	Seal Oil
Condition	6	4	24	Good condition
Existing Maintenance	5	4	20	Formal maintenance program in place
Complexity	3	3	9	Simple rotating machinary
TOTAL			247	* Total Weighted Score

5 RESULTS

Raw data was collected across a number of sites following the methodology outlined in Section 4. Three sites were selected as a pilot study series. The data displayed in Figure 2 represents the raw data collected during the application of the methodology. Raw data collected included:

- Asset identification;
- Score for each asset against each of the six criteria;
- Description against each criterion – this was defaulted to the scoring guide description but over written as appropriate.

The results of three sites are presented and compared in this section. These sites were:

1. Site 1 – Food Processing Plant - 454 Assets Scored
2. Site 2 – Brewery – 333 Assets Scored
3. Site 3 – Petrochemical Plant – 217 Assets Scored

Summarised results of the application of the methodology include:

- Numbers of safety critical assets (safety / environmental score greater than or equal to 7 indicating potential fatality or reportable environmental consequence);
- Number of output critical assets (Output score greater than 5 indicating a degree of loss of output or loss of service);
- Top 10 Critical assets for each site;
- Histogram of total score relative frequency of total weighted score for each site;
- Histogram of scores relating to output for each site.

All results presented were calculated with Safety / Environmental, Output, Severity, Condition, Existing Maintenance and Complexity weightings of 10, 8, 7, 6, 5 and 3 respectively.

5.1 Site 1 – Food Processing Plant

454 assets in total were scored for Site 1. 31% of assets were identified as safety critical with a Safety / Environmental score of 7 or over. 23% of assets were identified as output critical with an Output score of 5 or over. The top ten highest total weighted scores are shown in Table 4.

Table 4- Top 10 Critical Assets for Site 1

	UNIT	*ASSET*	*TWS*
1	PRODUCT FILL/PACK	FILLER	323
2	PRODUCT BLENDING	SCRAPE SURFACE HEAT EXCHANGER	322
3	STEAM	SERVICES BRIDGE (ADJ. PROD A)	310
4	LIQUID PACKAGING	LIQUID FILLER	294
5	PREPARATION	MIXER #2	284
6	PREPARATION	MIXER #1	284
7	COOKING	RETORT COOKERS #1 - #16	280
8	PRODUCT C	KETTLE	278
9	PRODUCT D	THERMOFORMING MACHINE	277
10	PRODUCT D	KETTLE	277

The relative frequency of total weighted scores is shown in Figure 2.

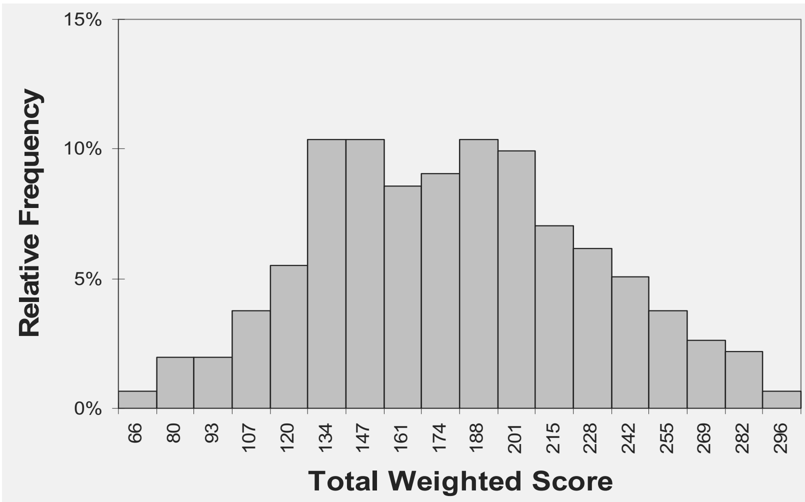


Figure 2- Relative Frequency of Total Weighted Scores for Site 1

The relative frequency of total output scores is shown in Figure 3

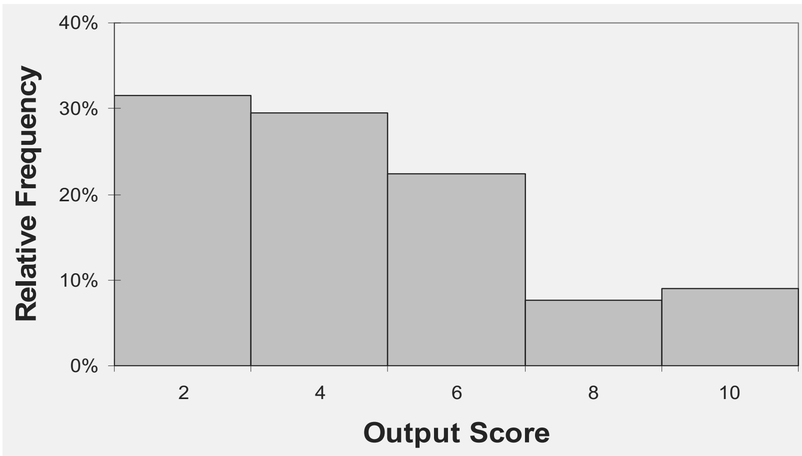


Figure 3- Relative Frequency of Output Scores for Site 1

5.2 Site 2 – Brewery

333 assets in total were scored for Site 2. 26% of assets were identified as safety critical with a Safety / Environmental score of 7 or over. 24% of assets were identified as output critical with an Output score of 5 or over. The top ten highest total weighted scores are shown in Table 5.

Table 5- Top 10 Critical Assets for Site 2

	UNIT	ASSET	TWS
1	GENERAL	PIPE BRIDGE	292
2	CIP SYSTEM	CIP CAUSTIC STORAGE	246
3	CIP SYSTEM	WORT COOLER CIP TANK	240
4	CIP SYSTEM	BREW VESSELS CIP TANK	240
5	BOTTLING PLANT	BOTTLE WASHER	237
6	BOTTLING PLANT	NO 1 FILLER	236
7	CANNING	CAN CRUSHER	235
8	BOTTLING PLANT	CIP SYSTEM	233
9	BOTTLING PLANT	# 1 FILLER	233
10	BOTTLING PLANT	CAUSTIC RECLAMATION	231

The relative frequency of total weighted scores is shown in Figure 4

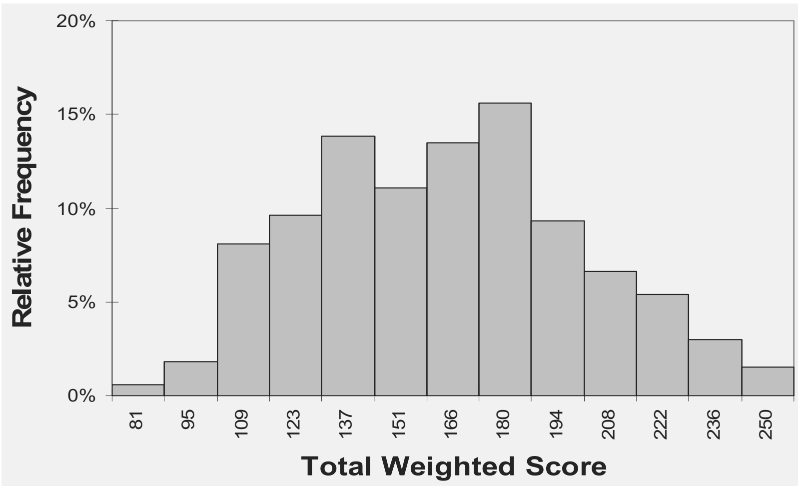


Figure 4- Relative Frequency of Total Weighted Scores for Site 2

The relative frequency of total output scores is shown in Figure 5

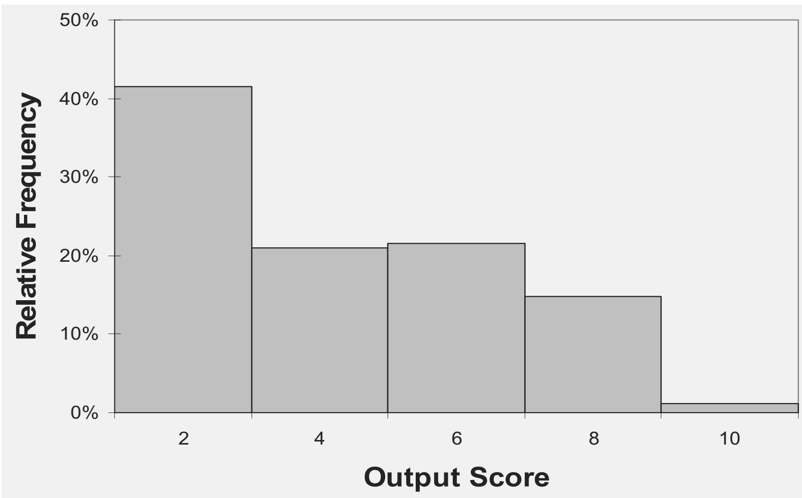


Figure 5- Relative Frequency of Total Weighted Scores for Site 2

5.3 Site 3 – Petrochemical Plant

217 assets in total were scored for Site 3. 10% of assets were identified as safety critical with a Safety / Environmental score of 7 or over. 13% of assets were identified as output critical with an Output score of 5 or over. The top ten highest total weighted scores are shown in Table 6.

Table 6- Top 10 Critical Assets for Site 3

	UNIT	ASSET	TWS
1	WHARF	FIRE FIGHTING SYSTEM	338
2	PLATINUM REFORMER	RECYCLE GAS COMPRESSOR	306
3	PLATINUM REFORMER	PTR REACTORS	284
4	WHARF	BLACK OIL CIRC PUMP	262
5	FLARE	MAIN FLARE HEADER	260
6	SULPHUR RECOVERY	CONDENSERS	250
7	SULPHUR RECOVERY	REHEATERS	250
8	SULPHUR RECOVERY	REACTOR	236
9	VACUUM DISTILLATION	CHARGE FURNACE	232
10	FUELS LOAD OUT POINTS	LPG LOAD OUT	224

The relative frequency of total weighted scores is shown in Figure 5

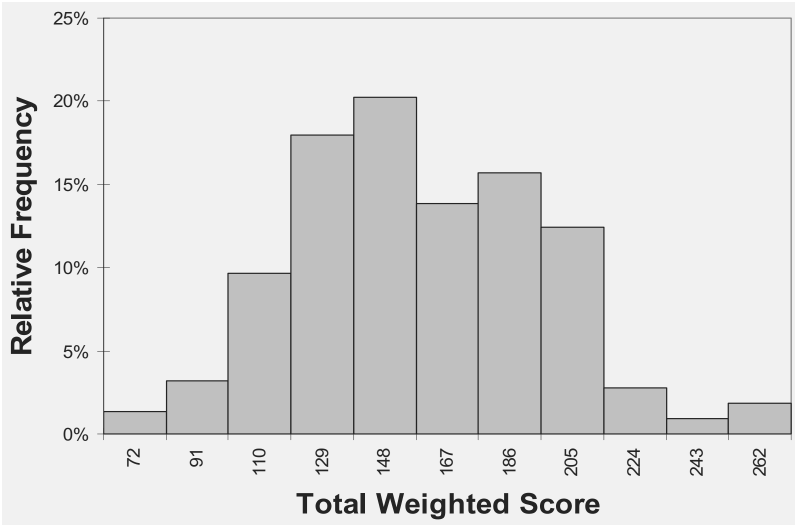


Figure 6- Relative Frequency of Total Weighted Scores for Site 3

The relative frequency of total output scores is shown in Figure 6

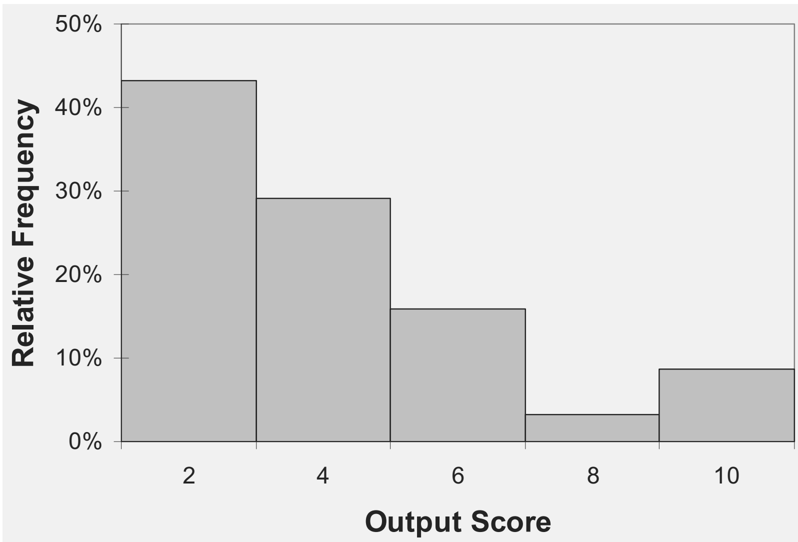


Figure 7 - Relative Frequency of Total Weighted Scores for Site 3

5.4 Discussion of Results

There are similarities between the results across the three sites studied. The following observations are made based on the results:

1. The results from the application of the methodology were well received by both the participants and other stakeholders. In the majority of cases, disagreements with the final ranking were not upheld after closer scrutiny of the criteria scoring.

2. There is variability in the Total Weighted Scores across the three sites supporting validity in the assumption of diversity of asset criticality across a facility.
3. The top ranked critical assets were dominated mainly by safety /environmental consequences. Although there were relatively high numbers of assets identified with safety / environmental consequences it was revealed that he number of maintainable items within the asset with a safety functionality were low.
4. Assets associated with the provision of facility services (air, water, steam etc.) featured prominently in the critical asset ranking. This was counter intuitive for many of the participants who expected the largest production assets to be the most critical.
5. There was a higher than expected percentage of assets assessed as having no measurable impact on the output of a facility. This was consistently demonstrated across all sites in the study series.

There were a number of lessons learned from the application of the methodology across the study series. This information was provided by participants in the application or subsequently by stakeholders or other observers. These lessons provided opportunities to modify the approach and apply additional work. These issues were:

1. Critical assets can disappear in the ranking. Assets with a high safety / environmental criticality with a low output criticality (or vice versa) could be significantly down the ranking. Separate lists should be maintained to segregate critical and non critical assets according to the ranking of the Total Weighted Score.
2. There is a need to improve the definition of an asset. Identification of assets is arbitrary and this method is not well described or defined.
3. Current criteria do not reflect the vulnerability of an asset to accidental damage or acts of vandalism. The criteria should be modified for subsequent applications.
4. There is an opportunity to document the methodology so that it can be successfully applied by others.
5. Computerisation of the data collection process greatly improved the efficiency and effectiveness of the application. Improvements to the current database would greatly enhance future applications.

Additional future work is warranted to satisfy these opportunities.

6 CONCLUSION

A systematic and standardized methodology was developed to identify critical assets. It was demonstrated in application that this methodology could be applied effectively and efficiently using the expert knowledge of those directly engaged in operations and maintenance functions. An assumption of diversity in asset criticality was supported by the application of this methodology. A variety in both ranking and relative criticality of assets was demonstrated consistently across the facilities included in the study series. The application of the developed methodology identified opportunities for future steps in the further development of expert knowledge based assessment tools of asset criticality.

7 REFERENCES

[1] Author Unknown, *A world leader in mining*, viewed 22 October 2005, http://www.riotinto.com/about/default.asp

[2] Hutton C, *Corporate Citizenship*, viewed 22 October 2005, http://www.surrey.ac.uk/Education/ETGACE/ukworkshop-ChrisH.doc

[3] Murray M et. al., *Maintenance Engineering Society of Australia Capability Assurance: A Generic Model of Maintenance*, viewed 12 February 2006, http://www.mesa.org.au/engineering.html

[4] Kelly A, *Maintenance Strategy - Business-Centered Maintenance*, Butterworth Heinemann, Oxford, 1997

[5] Moteff J Parfomak P, *CRS Report for Congress, Critical Infrastructure and Key Assets: Definition and Identification*, Congressional Research Service, The Library of Congress, 2004

[6] *International Infrastructure Management Manual, Australia/New Zealand Edition*, October 2002, Association of Local Government Engineering NZ Inc, Page xiii

[7] Dreher, L.G.; Jarman, M.F., *Application of 'life cycle' concept for asset risk management, International Conference on Health, Safety and Environment in Oil and Gas Exploration and Production*, March, 2004, SPE International Conference on Health, Safety and Environment in Oil and Gas Exploration and Production, p 521-524

[8] Jay P, *Joined Up Asset Management – Managing Priorities*, 71st Annual International Conference of Doble Clients, viewed 30 May 2006, http://www.twpl.co.uk/resources/criticality 20paper 202004 20rev 201.pdf

[9] Standards Association of Australia 2006, *Dependability management - Part 3-14: Application guide - Maintenance and maintenance support*, (IEC 60300-3-14 (2004-07), Standards Australia, North Sydney, Page 15

[10] Kelly A, *Maintenance and its management*, Conference Communication, 1989

[11] Author Unknown, *Asset Criticality & Business Impact Analysis*, Brunei Shell Petroleum, 1997, viewed 30 May 2006, http://www.brunet.bn/gov/y2k/bsc/ascritm.htm

[12] *Practices for Protecting Information Resources Assets. Identifying Critical Assets and Conducting a Risk Analysis and Assessment*, State of Texas Department of Information Resources , viewed 30 May 2006, http://www.dir.state.tx.us/IRAPC/practices/ , September 2003, Page 2.5

[13] IEC*, Reliability and Maintainability, CEI ICE 812, Analysis techniques for system reliability – Procedure for Failure Modes and Effect Analysis (FMEA)*, Bureau Central de la Commission Electronique Internationale, First Edition 1985

PROFESSIONAL SKILLS TRAINING IN INTEGRATED ASSET MANAGEMENT: HOW TO DEVELOP AND IMPLEMENT THE ESSENTIAL ORGANISATIONAL ASSET MANAGEMENT FUNCTIONS

RF Stapelberg

Education and Training Manager The Cooperative Research Centre for Integrated Engineering Asset Management - Australia

Abstract: Asset management in the context of physical infrastructure and industrial assets is the process of organising, planning and controlling the acquisition, use, care, refurbishment, and/or disposal of an organisation's physical assets to optimise their service delivery potential and to minimise the related risks and costs over their *entire life*. Asset management on national and international scales is extremely wide and complex, and consequently has either fallen short of specific requirements, or lacks application of the essential principles. There is therefore an emerging need for a significant increase in the level of competency and technological sophistication through *integration* of asset management decision-making across all public and private sectors. This competency includes the concept of *knowledge management decision-making* in integrated asset management, with due consideration of the essential differences of approach by the different public and private sectors; yet with a standardized implementation of education and training in integrated asset management that would have the widest reach and applicability on a national (and international) scale for each sector. This paper specifically considers standardized professional skills training in integrated asset management with particular reference to meeting industry needs of how to develop and implement the essential organizational asset management principles, themes, frameworks and functions, and to establish an *integrated knowledge-based framework* in which *knowledge-based decision-making* can be optimized.

Key Words: Infrastructure and industrial assets, asset management, integrated asset management, knowledge management decision-making, integrated knowledge-based frameworks, professional skills training.

1 INTRODUCTION

Integrated Asset Management (IAM) involves integrated processes of managing physical assets during their useful lives, and requires a certain level of management insight and expertise from diverse organisational disciplines. IAM is a systematic, structured process covering the whole life of physical assets whereby the underlying assumption is that an organisation's assets exist to support the organisation's delivery strategies. The principal objective of IAM is to achieve the best possible match of assets with an organisation's delivery strategies to achieve the desired outputs and outcomes. This is done through the appropriate development and implementation of the essential organizational asset management principles, themes, frameworks and functions as researched in current literature of the status of asset management world-wide [1].

2 ASSET MANAGEMENT PRINCIPLES, THEMES, FRAMEWORK, AND FUNCTIONS

The key starting point, to ensure assets support an organisation's delivery strategy, is to establish a link between delivery objectives and the assets. Corporate objectives are then translated into delivery strategies, outputs and outcomes. An organisation's delivery strategies consequently combine with information systems, personnel, and financial resources. Asset management strategy on the other hand, is not simply a summation of the individual plans developed for each phase of an asset's life-cycle. It must be consistent with corporate objectives and integrated with other management strategies [2].

It should therefore relate to specific organizational requirements that will ensure the correct establishment of certain asset management *principles*. The general concepts or *themes* of asset management in turn must relate to these *principles*. Furthermore, asset management decisions should not be made in isolation. They should be part of an overall *framework* of decision-making in an organisation. Asset management objectives can thus be translated into principles, themes, and frameworks. Once the asset management objectives have been translated into principles, themes, and frameworks, a concept map for the development of asset management *functions* can be developed in accordance with these objectives. A schematic presentation of such a concept map is indicated in Figure1:

Concept Map Stages of Developing Asset Management Functions:

Identify and generalise asset management ***principles***

⬇

Establish the relevant asset management ***themes***

⬇

Develop a comprehensive integrated asset management ***framework***

⬇

Initiate a first attempt at building a theoretical framework for asset management ***functions***

Figure 1. Concept Map of Developing Asset Management Functions [1]

2.1 Asset Management Principles

The *principles* of asset management common to most organisations are [1]:

- Asset management is driven by corporate/business objectives and goals:
- Assets exist to meet some identifiable service delivery. Asset management provides clearly assigned managerial accountabilities and responsibilities for service delivery:
- Asset management places emphasis on managing assets (what, why, where, how much):
- Asset management provides effective recognition and life cycle management of risks:
- Asset management places emphasis on optimising existing asset utilisation and performance to maintain or reduce overall service delivery cost to a sustainable level:
- Asset management considers the use of engineering solutions and management techniques for organising, planning and controlling the acquisition, use, care, refurbishment, and/or disposal of an organisation's physical assets.
- Asset management takes a total life-cycle cost approach and utilises “whole of life cycle” techniques for assets to optimise their service delivery potential and to minimise the related risks and costs over their *entire life*:
- Asset management focuses on performance measurement, monitoring and matching assets to organisational delivery strategies:

The *general principles* of asset management can thus be summarised as:

- An effective framework can be established for asset management;
- Asset management decisions are integrated with strategic planning;
- Asset planning decisions are based on the 'life-cycle' of assets;
- Asset ownership/stewardship bears risks that are assessed and evaluated;
- The appropriate tactical planning considers asset condition and use;
- Financial accountability is established for asset performance and residual life.

These principles are not definitive but are consistent with trends in improving asset management in the public and private sectors. Each principle reflects a fundamental notion of best practice and continual improvement, although they are not widely practiced in most industry sectors.

2.2 Asset Management Themes

The basic *themes* of asset management represent current thinking with national asset owners as well as professional organisations representing asset owners [1]. Underlying these themes is the assumption that assets exist to support program delivery, and relate to the following:

- Asset strategic planning;
- Asset ownership/stewardship;
- Asset service delivery and risk;
- Asset life cycle costing and budgeting;
- Asset usage, maintenance, and performance;
- Asset information and decision-making.

2.3 Asset Management Frameworks

So far, various asset management principles have been considered from the perspective of their related themes, and these themes have been used as a basis for comparison – which in effect can be presented in the form of an asset management structure or *framework*. Such a framework would include the following [1]:

- Strategic planning and decision-making;
- Ownership/stewardship and governance;
- Service delivery and risk analysis;
- Life cycle costing and budgeting;
- Engineering and economic analysis;
- Data management and information;
- Condition monitoring and modelling;
- Usage life-cycle and maintenance;
- Performance measure and financial management.

It is the various themes to which the general principles of asset management relate that can be presented in the form of an asset management structure or framework.

A study of various asset management frameworks developed by the public and private sector for both infrastructure and industrial assets revealed a fairly consistent adherence to the general asset management principles and themes. Although asset management requires a *framework* to support both short and long term organisational *delivery strategies*, the actual integration process was not explicitly defined up until the formalisation of an Integrated Asset Management Framework by the Cooperative Research Centre for Integrated Engineering Asset Management [3]. This integration process was reviewed by CIEAM from an adaptation of a Strategic and Tactical Asset Management framework developed by Stapelberg, (2000).

2.4 Integrated Asset Management Frameworks

Such an integrated framework includes the following modules [1]:

- Strategic planning of assets;
- Asset management decisions;
- Asset ownership/stewardship;
- Asset service delivery and risk analysis;
- Asset life cycle costing and budgeting;
- Asset data management;
- Asset condition monitoring;
- Asset engineering and economic analysis;
- Operating and maintaining assets;

- Asset usage information and feedback;
- Performance measure of assets;
- Assets financial management.

The Integrated Asset Management (IAM) Framework was adapted from the asset management integration approach developed by Stapelberg (2000) which is divided into five areas:

- Strategic planning and risk integration.
- Budgeting and costing and financial management integration.
- Data management and information systems integration.
- Condition monitoring and performance measure integration.
- Tactical planning and asset usage life-cycle integration.

The approach considers a 'top-down/bottom-up' systems integration from key strategies and higher order asset management needs, through the appropriate budgeting and costing and financial management integration, to data management and information systems integration. It is specifically the data management and information systems level that forms the bridge between strategic planning and tactical planning, whereby the bottom-up approach progresses from tactical planning and asset usage life-cycle integration, through condition monitoring and performance measure, to data management and information systems integration. This 'top-down/bottom-up' integrated process appears linear, although it is in effect cyclic and follows the asset life-cycle from conceptualization through establishment to disposal.

The Stapelberg (2000) asset management model shows an effective asset management framework of ten core modules or functions which incorporate an asset management strategy integration concept and an asset management tactics integration concept that optimizes asset value over the complete asset life cycle. This is illustrated in Figure 2. This model divides the asset management system integration approach into two main concepts namely:

The strategy integration concept: The objective of this concept is to provide a strategic planning model that incorporates all the higher order asset management systems within the organization such as the overall asset management plans, assets risk management, budgeting and costing and assets financial management. This provides a consistent view of the organization's *strategic business model* across all asset management systems and serves as a basis for the integration of these systems. The products of this approach include high level business strategies.

The tactics integration concept: This concept identifies the *operational functions* of each system in the entire asset management program. It helps to establish the integration requirements of all systems at their process level. The tactics integration concept also considers three distinct integration phases: a Data Management - Information Systems Integration Phase; a Condition Monitoring - Performance Measures Integration Phase; and Tactical Planning - Assets Usage Life-Cycle Integration Phase.

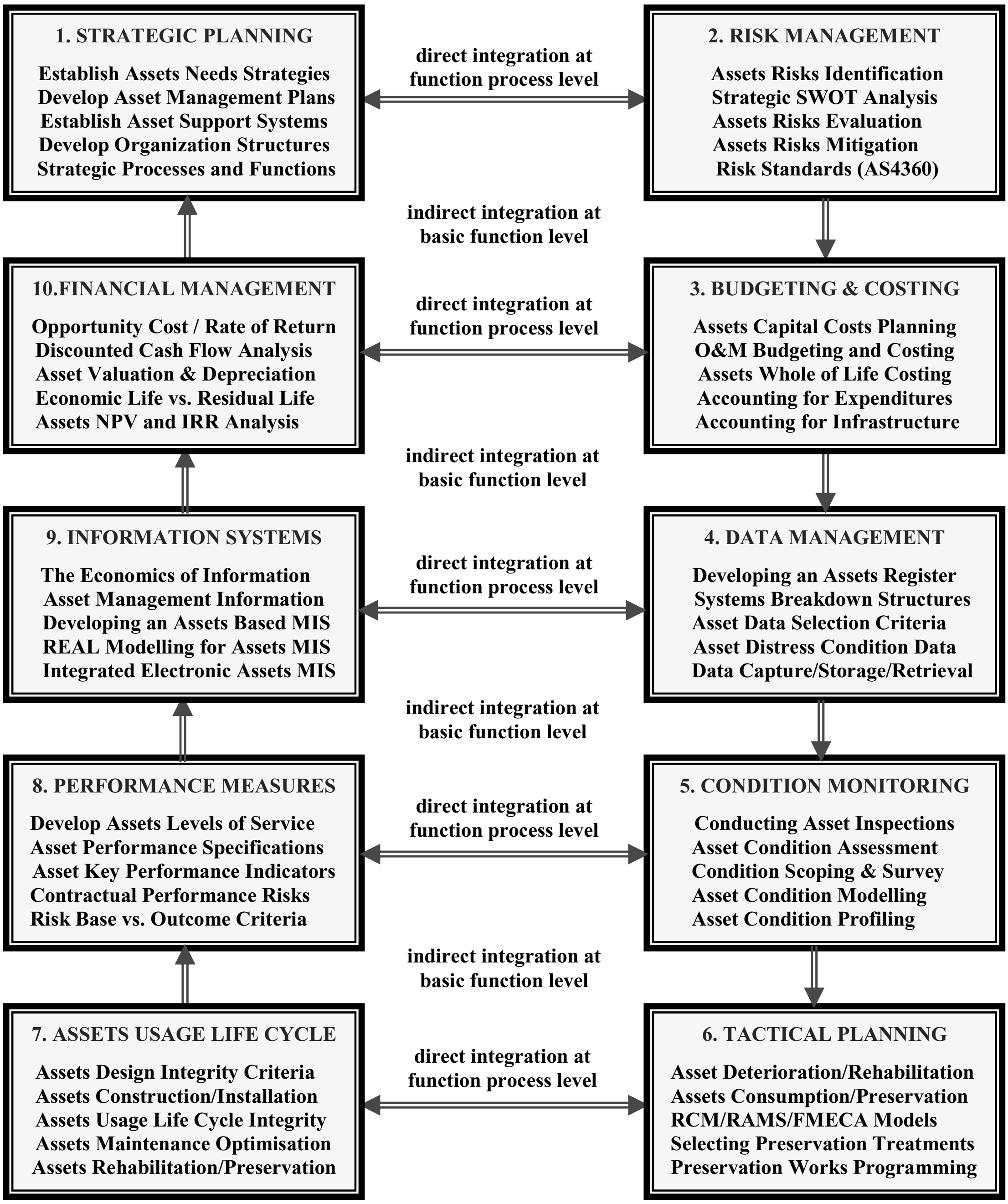


Figure 2. Integrated Asset Management Integration Concept [4]

The Stapelberg (2000) asset management model seeks to integrate an organization's *strategic business model* across all asset management systems with the *operational functions* of each of these systems through a process of *knowledge management decision-making*. The model provides an effective framework for the development of various functions and sub-functions of each core module, as well as the integration of these sub-functions through related operational processes and appropriate decision-making.

3 ASSET MANAGEMENT FUNCTIONS

The development of various ***functions*** of each core module of an Integrated Asset Management Framework as developed by the Cooperative Research Centre for Integrated Engineering Asset Management [3], and adapted from a Strategic and Tactical Asset Management framework developed by Stapelberg, [4], includes the following:

- Strategic planning;
- Asset ownership;
- Risk management;

- Budgeting and costing;
- Data management;
- Condition monitoring;
- Tactical planning;
- Human resources;
- Assets usage life-cycle;
- Performance measure;
- Information systems;
- Financial management.

The building of a theoretical framework for ***asset management functions and sub-functions*** is presented in modular format in Figure 3 [1].

1. STRATEGIC PLANNING **Establish Assets Needs Strategies** **Develop Asset Management Plans** **Establish Asset Support Systems** **Assets/Corporate Purpose Links** **Strategic Processes and Functions**	2. ASSET OWNERSHIP **Assets Ownership/Stewardship** **Ownership Responsibilities** **Strategic Governance** **Stakeholder Relationships** **Assets Alliances**	3. RISK MANAGEMENT **Assets Risks Identification** **Strategic SWOT Analysis** **Assets Risks Evaluation** **Assets Risk Mitigation** **Risk Standards (AS4360)**
4. BUDGETING & COSTI NG **Assets Capital Costs Planning** **O&M Budgeting and Costing** **Assets Whole of Life Costing** **Accounting for Expenditures** **Accounting for Technology**	5. DATA MANAGEMENT **Developing an Assets Register** **Systems Breakdown Structures** **Asset Data Selection Criteria** **Asset Distress Condition Data** **Data Capture/Storage/Retrieval**	6. CONDITION MONITORING **Conducting Asset Inspections** **Asset Condition Assessment** **Condition Scoping & Survey** **Asset Condition Modelling** **Asset Condition Profiling**
7. TACTICAL PLANNING **Asset Deterioration/Rehabilitation** **Assets Consumption/Preservation** **Assets Diagnostics and Prognosis** **RCM/RAMS/FMECA Modelling** **Selecting Preservation Treatments**	8. HUMAN RESOURCES **Assets Education and Training** **HR Strategic Planning** **Assets Systems Utilisation** **Employment Relations** **Teams & Virtual Teams**	9. ASSETS USAGE LIFE CYCLE **Assets Design Integrity Criteria** **Assets Construction/Installation** **Assets Usage Life Cycle Integrity** **Assets Maintenance Optimisation** **Assets End of Usage / Disposal**
10. PERFORMANCE MEASURE **Develop Assets Levels of Service** **Asset Performance Specifications** **Asset Key Performance Indicators** **Contractual Performance** **Risk Base vs. Outcome Criteria**	11. INFORMATION SYSTEMS **The Economics of Information** **MI Access/Validity/Dispersion** **Developing an Assets Based MIS** **REAL Modelling for Assets MIS** **Integrated Electronic Assets MIS**	12. FINANCIAL MANAGEMENT **Funding / Discounted Cash Flow** **Basis for Asset Valuation** **Assets Opportunity Cost / ROI** **Economic Life vs. Residual Life** **Assets Depreciation / Disposal**

Figure 3. Diagrammatic Structure of a Theoretical Framework [1]

This framework incorporates core modules or functions, as well as their sub-functions. These are listed as:

- **Strategic planning** – Establish assets needs strategies; develop asset management plans; asset support systems; asset/corporate purpose links; strategic processes and functions.
- **Asset ownership** – Asset ownership/stewardship; ownership responsibilities; strategic governance; stakeholder relationships; assets alliances.
- **Risk management** – Assets risks identification; strategic SWOT analysis; assets risks evaluation; assets risk mitigation; risk standards (AS4360).
- **Budgeting and costing** – Assets capital costs planning; O&M budgeting and costing; assets whole-of-life costing; accounting for expenditures; accounting for technology.

- **Data management** – Developing an assets register; systems breakdown structures; asset data selection criteria; asset distress condition data; data capture/storage/retrieval.
- **Condition monitoring** – Conducting asset inspections; asset condition assessment; condition scoping and survey; asset condition modeling; asset condition profiling.
- **Tactical planning** – Asset deterioration/rehabilitation; assets consumption/preservation; assets diagnostics and prognosis; reliability modeling; selecting preservation treatments.
- **Human resources** – Assets education and training; HR strategic planning; assets systems utilization; employment relations; teams and virtual teams.
- **Asset usage life-cycle** – Assets design integrity criteria; assets construction/installation; assets usage life-cycle integrity; maintenance optimization; assets end-of-use/disposal
- **Performance measure** – Develop Level-Of-Service; asset performance specifications; key performance indicators; contractual performance; risk base versus outcome criteria.
- **Information systems** – Economics of information; MI access/validity/dispersion; developing an assets MIS; REAL modeling for assets MIS; integrated electronic MIS.
- **Financial management** – Assets funding/DCF; assets valuation; assets opportunity costs/ROI; economic life versus residual life; assets depreciation/disposal.

It is in accordance with these functions that a schema can be developed for professional skills training in Integrated Asset Management. However, in addition to the application of these functions and sub-functions of asset management in a structure or framework, it is essential to determine whether they are indeed applied with the necessary knowledge and insight through an integration of the various organisational disciplines involved in asset management. From such an approach, the fundamentally unique concept of **Integrated Knowledge-based Asset Management** can be developed [5].

4 INTEGRATED KNOWLEDGE-BASED ASSET MANAGEMENT

Integrated Knowledge-based Asset Management involves *integrated* processes and disciplines for managing organisational physical assets, and requires a certain level of *knowledge management* and *decision-making insight* from diverse organisational functions. It is therefore a systematic, structured process covering the whole life of an organisation's physical assets, whereby the underlying assumption is that the concept of *knowledge-based decision-making* in regard to these assets forms the basis of the organisation's delivery strategies. The principal decision objective of Integrated Knowledge-based Asset Management is thus to achieve the best possible match of assets within an organisation's delivery strategies. This could however pose an essential reliance upon an *integrated knowledge framework for decision-making* in the implementation of such delivery strategies, additionally to establishing an effective integrated framework for asset management. The following application of *knowledge management* and *decision-making* to the principles of asset management thus represents a *new thinking paradigm* amongst asset owners and asset management organisations for the implementation of Integrated Asset Management.

4.1 Integrated Knowledge-based Asset Management Principles

The principles of integrated knowledge-based asset management can be summarized as [5]:

- A total *integrated knowledge framework* can be established for asset management;
- Knowledge management decision-making is integrated in asset *strategic planning*;
- Strategic planning decisions are based on *life-cycle knowledge* of physical assets;
- Asset ownership and/or stewardship bears *risks* that can be assessed and evaluated;
- Asset tactical planning considers *specific knowledge* of asset conditions and safety;
- Financial *accountability* is based on asset performance and residual life knowledge.

Each principle now reflects a fundamental notion of dependency upon knowledge management and decision-making in IAM in the public and private sectors.

4.2 Knowledge Management in Integrated Asset Management

Knowledge management (KM) is concerned with the exploitation and development of the knowledge assets of an organisation with a view to furthering the organisation's objectives. The knowledge to be managed includes both explicit, documented knowledge, and tacit, subjective knowledge. Organisations that succeed in knowledge management are likely to view knowledge as an asset and to develop organisational norms and values, which support the creation, and sharing of knowledge [6]. Historically, KM has been aimed at managers through what has been generally referred to as an executive information system aimed at supporting managerial decision-making. More recently, however, KM systems are increasingly designed for entire organisations. As organisations are being driven toward KM to meet organisational delivery strategies and create value, they are increasingly finding that use of existing knowledge and the creation of new knowledge inevitably allows for better decision-making within an integrated asset management framework. In addition, KM technology is ideally suited for organisational groups such as Asset Management Groups (AMG's).

5 CONCLUSION: A SCHEMA FOR PROFESSIONAL SKILLS TRAINING IN IAM

With the development of the essential functions and sub-functions that can be applied with the necessary *knowledge* and *insight* through an *integration* of the various organisational disciplines involved in asset management, a schema can now be developed for professional skills training in IAM. Such a schema is referenced in Figure 4 [4].

6 REFERENCES

[1] Stapelberg, R.F., "Australian Infrastructure and Industry Assets Management Survey: Preliminary Literature Review and Survey Analysis Report", CIEAM Research Report ET01/0106, the Cooperative Research Centre for Integrated Engineering Asset Management, January 2006.

[2] CAM, "Commonwealth Asset Management Handbook - Better Practice Guide" Commonwealth Auditor-General, Audit Report No. 27 of 1995-96, Financial Control and Administration Audit on Asset Management, June 1996.

[3] CIEAM, "Integrated Engineering Asset Management", www.cieam.com.au, The Cooperative Research Centre for Integrated Engineering Asset Management, O-Block, Queensland University of Technology, Gardens Point, Brisbane Australia, 2003.

[4] Stapelberg, R.F., "Strategic and Tactical Asset Management", Asset Management Training Manual, ICS Industrial Consulting Services Pty Ltd, Miami QLD, Australia, 2000.

[5] Stapelberg, R.F. "Research into the Development and Application of Integrated Asset Management in the Australian Public and Private Sectors", unpublished paper from DBA thesis, Graduate College of Management, Southern Cross University, 2005.

[6] Rowley, J. (2000), "From learning organisation to knowledge entrepreneur". Journal of Knowledge Management, 4, 1, 2000, 7-14.

Assets Strategic Planning
- Establishing a Needs Strategy for Various Types of Assets.
- Developing Assets Management Plans for ISO Accreditation.
- Establishing Asset Support Systems for Effective Assets Management.
- Developing Assets Management Functional Organization Structures.
- Developing Strategic Assets Management Core Processes and Functions.

Assets Risk Management
- Identifying Asset-based Risks for Various Types of Assets.
- Conducting Strategic and Tactical SWOT Analysis.
- Evaluating Asset-based Risks for Various Types of Assets.
- Identifying Risk Mitigation Methods for Various Types of Assets.
- Applying the Australian Risk Management Standard AS4360.

Assets Budgeting and Costing
- Planning for Assets Capital Costs.
- Budgeting and Costing for Assets Operations and Maintenance.
- Using Whole-of-Life Costing Methods for Various Types of Assets.
- Accounting for Asset Management Expenditures.
- Accounting for Infrastructure Assets.

Assets Data Management
- Developing an Assets Register with the Relevant Assets Data.
- Developing Systems Breakdown Structures for Various Types of Assets.
- Establishing Criteria for Critical and Non-Critical Asset Data Selection.
- Collecting Asset Distress Condition Data for Specific Types of Assets.
- Using Mechanical and Electronic Data Capture, Storage and Retrieval.

Assets Condition Monitoring
- Conducting Defect and Condition-based Asset Inspections and Surveillance.
- Establishing Condition Assessment Techniques for Various Types of Assets.
- Applying In-depth Asset Condition Scoping and Asset Survey Techniques.
- Using Asset Condition Modelling Techniques to Determine Asset Condition.
- Developing Asset Condition Profiles of Different Asset Distress Conditions.

Assets Tactical Planning
- Determining Assets Deterioration and Rehabilitation Criteria.
- Evaluating Assets Consumption versus Preservation Threshold.
- Applying PMS/RAMS/FMECA Models for Assets Preservation.
- Selecting Effective Asset Preservation Treatments at Budgeted Costs.
- Programming Cost-effective Asset Maintenance/Rehabilitation Works.

Assets Usage Life Cycle
- Evaluating Assets Design Integrity Criteria for Various Types of Assets.
- Evaluating Assets Construction/Installation/Operation Integrity Criteria.
- Evaluating Assets Usage Life Cycle Integrity for Various Types of Assets.
- Optimising Assets Maintenance Using Different Types of Maintenance.
- Evaluating Assets Maintenance/Rehabilitation/Preservation Strategies.

Assets Performance Measures
- Developing an Assets Level of Service (LOS) Performance Measure.
- Establishing and Understanding Asset Performance Specifications.
- Developing and Implementing Asset Key Performance Indicators.
- Identifying/Evaluating Contractual Asset Management Performance Risks.
- Evaluating Risk-base versus Outcome Contractual Performance Criteria.

Assets Information Systems
- Understanding the Economics of Asset Management Information.
- Evaluating Management Information Systems for Various Types of Assets.
- Developing an Integrated Assets-based Management Information System.
- Developing a REAL Model for an Assets Management Information System.
- Implementing Integrated Electronic Assets Management Information Systems.

Assets Financial Management
- Analysing Opportunity Cost versus Rate of Return in Assets Acquisition.
- Applying Discounted Cash Flow (DCF) Techniques in Asset Management.
- Understanding and Using Asset Valuation and Depreciation Methods.
- Understanding and Using Asset Economic Life versus Asset Residual Life.
- Determining Assets NPV and IRR for Assets Financial Management.

Figure 4. Schema for Professional Skills Training in IAM [4]

SENSOR DEGRADATION DETECTION IN LINEAR SYSTEMS

L JIANG [a] D Djurdjanovic [a] J Ni [a,c], J Lee[c]

[a] *Department of Mechanical Engineering, NSF I/UCRC on Intelligent Maintenance Systems (IMS), University of Michigan, 2350 Hayward Street, Ann Arbor, MI 48109-2125, USA*

[b] *Department of Mechanical, Industrial and Nuclear Engineering, NSF I/UCRC on Intelligent Maintenance Systems (IMS), University of Cincinnati, 2600 Clifton Avenue, Cincinnati, OH 45221, USA*

[c] *S.M. Wu Manufacturing Research Center, University of Michigan, Ann Arbor, MI, USA*

Abstract: As a result of aging and wear, the performance of machine and equipment deteriorates, which leads to a decrease in performance reliability and increase in the potential for faults and failures. To ensure proper functionality of complex systems, an ever-increasing number of increasingly sophisticated sensors and measurement devices have been employed for the purpose of performance diagnosis and control. Nevertheless, sensors themselves degrade, just as any other dynamic system. Therefore, to improve the reliability and performance of the system, it is essential to track and monitor the health conditions of the sensors as well. Without imposing requirements for redundant sensing equipment, the method developed in this paper utilizes the fact that sensor readings depict dynamic characteristics of the measured phenomenon in the plant as well as the dynamic characteristics of the sensor itself, which enables detection and quantification of sensor performance degradation as well as its isolation from degradation processes in the monitored or controlled plant. The method assesses sensor performance based on identification and tracing of time-constants and gains-changes associated with the sensor, which are obtained from using the subspace model based system identification method and that based on Auto-regressive Moving Average model with eXogeneous variables (ARMAX model), respectively. The method was verified through high-fidelity simulations of an automotive electronic throttle system.

Key Words: Sensor Performance Assessment, Subspace Model Based System Identification Method, ARMAX Modelling, Confidence Value (CV)

1 INTRODUCTION

The performance of machines and equipment degrades as a result of aging and wear, which decreases performance reliability and increases the potential for faults and failures [1]. To ensure proper functionality of complex systems, advanced technologies for performance diagnosis and control are incorporated into engineering designs, especially in the case of sophisticated, expensive and safety critical systems, such as manufacturing equipment, computer networks, automotive and aircraft engines, etc. Reliable performance diagnosis and control functionalities necessitate the use of an ever-increasing number of increasingly sophisticated sensors and measurement devices that are capable of delivering data about the key indicators of system status and performance. Nevertheless, sensors themselves degrade, just as any other dynamic system. Incipient failures such as bias, drift and precision degradation in the sensors can lead to errors in system condition diagnosis and ineffective control because the sensors are no longer able to deliver accurate information about the system's health condition. Therefore, to improve the reliability and performance of diagnostic and control systems, it is essential to ensure the health of the sensors as well, which places great importance on the problem of sensor performance monitoring.

The reliability of sensor readings can be improved by introducing redundant sensing hardware into the system. Such an approach is regularly used in safety-oriented applications such as aircraft and nuclear-power plant monitoring and control. However, the use of redundant sensing devices may not always be feasible due to cost and space constraints.

In this paper, we utilized the fact that the sensor readings depict dynamic characteristics of the measured phenomenon in the plant or system as well as the dynamic characteristics of the sensor itself. The aim of the method presented in this paper is to capture dynamic characteristics of the sensor within the sensor readings and separate them from those of the monitored phenomenon in the monitored or controlled system, thus enabling one to track the key dynamic indicators of the sensor and statistically assess the significance of their changes, without explicit need for redundant sensing hardware.

The reminder of this paper is organized as follows. A brief literature state-of-the-art review on the techniques on sensor reading validation is given in Section 2. A detail description of the sensor performance assessment method is presented in Section 3. To verify the developed method, several degradation scenarios of an automotive electronic throttle system in the simulation environment are conducted in Section 4. Finally, conclusions and possible future work of the current research work are given in Section 5.

2 STATE-OF-THE-ART REVIEW

Monitoring of sensor performance is becoming increasingly important as on-board system control, condition-diagnosis and performance prediction are being implemented in an ever-increasing number of dynamic systems. A conventional engineering method for ensuring sensor validity is to check and recalibrate the sensors periodically by following a set of predetermined procedures [2]. Although it is effective in dealing with bias problems, such method is not capable of continuously assessing the sensor performance and detecting small degradation and drifts in sensor performance. As more and more electronic components and sensors are being utilized, periodic checking of the sensors is no longer sufficient, especially in the case of sophisticated, expensive, and safety critical systems. Therefore, two alternatives are pursued in assessment of sensor performance: hardware redundancy approach and analytical redundancy approach [3].

The general idea of the hardware redundancy method is to measure the same key variables using two or more sensors. It is widely used in safety critical applications, such as aircrafts, submarines or space vehicles. Checking of the consistency of the redundant measurements of key process variables readily yields the faulty sensor, which can be picked out, for instance, by employing majority vote algorithm or by conducting the correlation analysis. Through the use of such a method, a monitoring system with higher measurement precision can be implemented using a group of sensors with comparatively lower precision [4]. Nevertheless, although the hardware redundancy method is effective and simple to implement, the redundant sensors increase the cost and impose larger space requirements on the entire system.

On the other hand, analytical redundancy approach builds up the static and dynamic relations between the multiple sensors based on a known or empirically obtained mathematical model of the corresponding system. Then, in order to detect the faulty sensor, residuals are generated by comparing the actual sensor readings and the readings estimated through the model. Fewer sensors are needed to pursue analytical redundancy than hardware redundancy. However, such method requires a significant amount of prior knowledge about sensor and plant dynamics and behaviour in order to detect sensor degradation and decouple it from degradation of the plant that is being measured.

In recent years, neural network based estimators are used to observe the critical sensor by taking the measurements of its neighbouring sensors as inputs. In [5], an approach is proposed to detect the defective sensor by checking whether its reading is far from the established value estimated by a Cerebellar Model Articulation Controller (CMAC) neural network from the measurements of its correlated sensors. In addition, the reading of the defective sensor can be reconstructed as the estimated value from the CMAC neural network. The problem of network architecture and appropriate training was not addressed. G. G. Yen *et al.* proposed a Winner Take All Experts (WTAE) network based on a ‘divide and conquer’ strategy, which significantly reduces the computational time required to train the neural network [6]. M. R. Napolitano *et al.* evaluated the robustness of the on-line learning neural networks and Kalman filters based approaches by applying both approaches for several different types of faults [7]. S. Wang *et al.* presented an online fault detection and diagnosis and sensor reconstruction scheme with an application of central chilling systems. The method is based on the Principal Component Analysis (PCA) to build the model that captures the correlation between the different sensors [8]. G. Kerschen *et al.* presented a procedure based on PCA to detect, isolate and reconstruct a faulty sensor for structural health monitoring [9]. Pablo *et al.* presented an algorithm for intelligent sensor validation in real time environment with a second Bayesian network that represents the dependencies and independencies between different sensors [3]. With such Bayesian network, the trade-off between the decision accuracy and time can be handled.

Nevertheless, the aforementioned work did not consider the possibility of the monitored plant degrading along with the sensor, and the subsequent need to decouple sensor dynamics from the remainder of the system in order to detect sensor dynamics degradation. The aforementioned problem will be addressed in this paper.

3 SENSOR PERFORMANCE ASSESSMENT METHOD

The problem of sensor performance assessment considered in this paper is based on the system structure shown in Figure 1, where sensed quantity contains the dynamics of the monitored or controlled system and the sensor, as well as the influence of process disturbances $w_p(t)$ and measurement noise $w_n(t)$. Sensor performance will be assessed using the observed control signal $u(t)$ and measured signal $y(t)$ without the use of redundant sensing equipment. The following assumptions will be introduced:

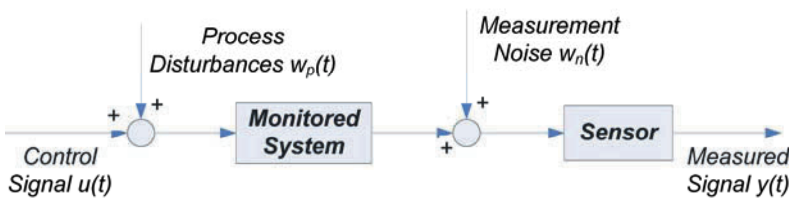


Figure 1. Structure of the compound system

a) There is no nonlinearity involved in the monitored system and the sensor.

b) The dynamics of the sensor is much faster than that of the monitored system. This assumption is not very restrictive because the dynamics of the sensor should be at least 2-5 times faster than that of the monitored system, which ensures accurate measurements and capturing of the dynamics of the monitored system.

c) The process disturbances and measurement noise are white noise.

Due to its linearity and causality of the compound system, the measured signal $y(t)$ cab be described in terms of the known input $u(t)$ and the unknown external disturbances $w_p(t)$ and $w_n(t)$ with the relationship in Equation (3.1).

$$y(s) = G_p G_s (u(s) + w_p(s)) + G_s w_n(s) \tag{3.1}$$

where G_p and G_s contain the dynamics of the monitored system and the sensor, respectively. Transfer functions $G_p(s)$ and $G_s(s)$ can then be represented as

$$\begin{aligned} G_p &= \frac{K'_p N_p(s)}{D_p(s)} = \frac{K_p(s+z_{p_1})(s+z_{p_1})\dots(s+z_{p_{n_p}})}{(\tau_1^p s+1)(\tau_2^p s+1)\dots(\tau_{m_p}^p s+1)}, n_p < m_p \\ G_p &= \frac{K'_s N_s(s)}{D_s(s)} = \frac{K_s(s+z_{s_1})(s+z_{s_1})\dots(s+z_{s_{n_s}})}{(\tau_1^s s+1)(\tau_2^s s+1)\dots(\tau_{m_s}^s s+1)}, n_s < m_s \end{aligned} \tag{3.2}$$

where K_p and K_s denote the gain of the monitored system and sensor, and τ^p and τ^s are the time constants associated with the poles of the monitored system and sensor, respectively.

Dynamic characteristics of the sensor can be generally described by the gain and the time constants of its transfer function poles, indicating the magnification capability and the response speed of the sensor. A sensor generally will respond slower or have deteriorated magnification capability as it degrades, which corresponds to the increasing time constant and changing (usually decreasing) sensor gain. Similarly, the performance of the monitored system can also be described by its gain (magnification capability) and time constant (speed). Thus, to identify the degraded sensor and distinguish the degradation source, the gain and time constants of the sensor and the plant should be tracked and monitored.

3.1 Detecting degradation in dynamic sensor parameters

Due to its superior performance in dealing with signals with dramatic changes, the subspace model based method is proposed to identify the linear compound structure of the sensor and the monitored system [9]. Based on the control signals and measured outputs, the subspace model based method estimates a state space model of the compound system in the form

$$\begin{aligned} x(t+T_s) &= Ax(t) + By(t) + Ke(t) \\ y(t) &= Cx(t) + Du(t) + e(t) \end{aligned} \tag{3.3}$$

where T_s denotes the sampling time, A, B, C and D are system matrices and K is the Kalman gain obtained by solving the corresponding Riccati Equation. The first line in Equation (3.3) is the state equation and the second is the output equation.

The subspace model based method provides an efficient and numerically reliable way to determine the predictors by projections directly on the observed data sequences [10]. With such predictors, all states $x(t)$ in Equation (3.3) can be obtained as a linear combination of the k-steps ahead predicted outputs (k = 1, 2, ……, n). Thus, if time sequences of state vectors $x(t)$, output vectors $y(t)$, and input vectors $u(t)$ are known, Equation (3.3) is actually a linear regression model and C and D can be estimated using the least square method, yielding error vector $e(t)$ from the output equation in (3.3). Given $e(t)$, the state equation in (3.3) becomes another linear regression in A, B, and K. (for more details, see Section 7.3 and 10.6 in [10] for details)

Transformation of the state space representation into a transfer function and tracking of the poles of the resulting transfer function readily yields the temporal behaviour of all plant and sensor poles and their corresponding time constants. Since the monitoring system should have much faster dynamic response than the monitored system in order to ensure rapid and accurate measurements, the time constants of the monitored system can be identified as much larger than that of the sensor. Thus, the degraded sensor with changes in its dynamics can be detected and its degradation can be distinguished from that of the monitored system.

Since the time constant of the degrading sensor changes gradually, the compound system is actually a time varying system which can be represented by a state space model with time varying matrices A, B, C, D, and K. To capture the time varying matrices, subspace model based method is used to fit a proper state space model to the data within a moving window. The flow chart shown in Figure 2 summarizes the procedures to detect and isolate the degradation in sensor dynamic parameters.

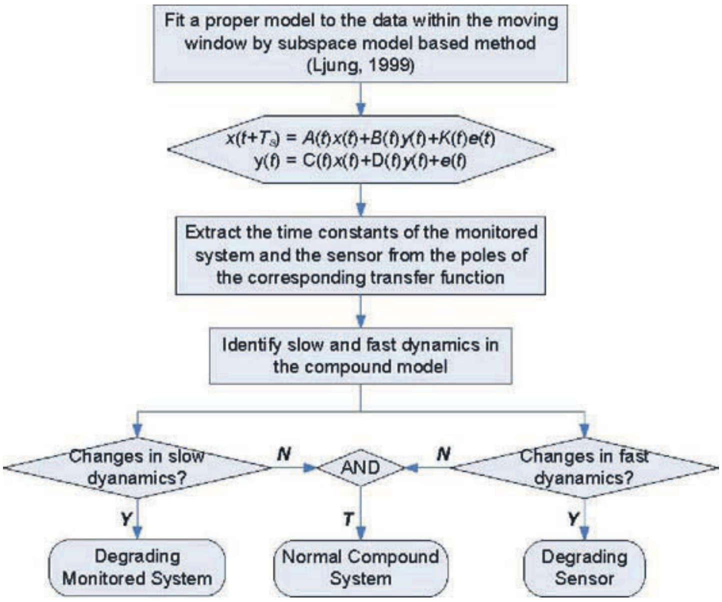


Figure 2. Flow chart of the procedures to detect degraded sensor with dynamic changes

3.2 Detecting degradation in sensor gain

The ARMAX model, a generalization of AutoRegressive Moving Average (ARMA) model, is able to incorporate an external input variable, thus establishing the relationship between the observed inputs and outputs. The form of the ARMAX model is

$$A(q)y(t) = B(q)u(t-nk) + C(q)e(t) \tag{3.4}$$

where *nk* denotes the input delay, and $u(t)$, $y(t)$ and $e(t)$ are the inputs, outputs and residuals, respectively. $A(q)$, $B(q)$ and $C(q)$, which represent the AutoRegressive (AR), eXogeneous (X) and Moving Average (MA) parts of an ARMAX model of order [*na*, *nb*, *nc*], can be expressed as

$$\begin{aligned} A(q) &= 1-\Phi_1 q^{-1} - \Phi_2 q^{-2} - \ldots - \Phi_{na} q^{-na} \\ B(q) &= \gamma_1 - \gamma_2 q^{-1} - \gamma_3 q^{-2} - \ldots - \gamma_{nb} q^{-nb+1} \\ C(q) &= 1-\Theta_1 q^{-1} - \Theta_2 q^{-2} - \ldots - \Theta_{nc} q^{-nc} \end{aligned} \tag{3.5}$$

As described in [11], the set of parameters of an ARMAX model is estimated by minimizing the sum of squared residuals. As the model becomes more complex, the associated sum of squared residuals decreases, however, the number of estimation parameters also increases. To achieve reasonable modeling accuracy with necessary complexity, either Akaike Information Criterion (AIC) or F-test criterion can be used to determine the order of the ARMAX model.

By comparing Equation (3.1) and (3.4), the relationship between the transfer of the compound system and the corresponding ARMAX model can be concluded as follows

$$\begin{aligned} A(q) &\leftrightarrow D_p(s)D_s(s) \\ B(q) &\leftrightarrow K_p^{'}K_s^{'}N_p(s)N_s(s) \\ C(q) &\leftrightarrow N_s(s)(K_p^{'}N_p(s)\frac{W_p(\omega)}{W_n(\omega)} + D_p(s)) \end{aligned} \tag{3.6}$$

where both sides of the operator ↔ can be derived from each other.

Consequently, by analogy, the AR part of the properly fitted ARMAX model between the control signal and the measured outputs contains the dynamics of the compound system.

$$e(t) = K_s' w_n(t) \Rightarrow std(e(t)) = K_s' std(w_n(t)) \tag{3.7}$$

$$K_{s_0}' = \frac{std(e_0(t))}{std(w_{n_0}(t))} \quad K_{s_k}' = \frac{std(e_k(t))}{std(w_{n_k}(t))} \Rightarrow K_s = \frac{K_{s_k}'}{K_{s_0}'} = \frac{std(e_k(t))}{std(e_0(t))} \bullet \frac{std(w_{n_0}(t))}{std(w_{n_k}(t))} = \frac{std(e_k(t))}{std(e_0(t))} \tag{3.8}$$

Furthermore, a proportional relationship in Equation (3.7) exists between the measurement noise $w_n(t)$ and the residuals of the adequate ARMAX model, which leads to their proportional statistical relationship. Thus, the sensor gain over a time window k and a reference window *0* can be defined according to Equation (3.8). Since the standard deviation of the white noise over any time window equals, the normalized sensor gains K_s over time can be obtained as the ratio between the standard deviations of the ARMAX model residuals over a time window and a reference window.

Based on the transfer function of the compound system identified by the subspace model based method, the product of the normalized gain of the sensor and the monitored system can be derived. Therefore, the normalized gain of the monitored system can be tracked, which enables the detection and isolation of the gain degradation in the compound system. Figure 3 summarizes the procedures to detect the degrading sensor with gain changes.

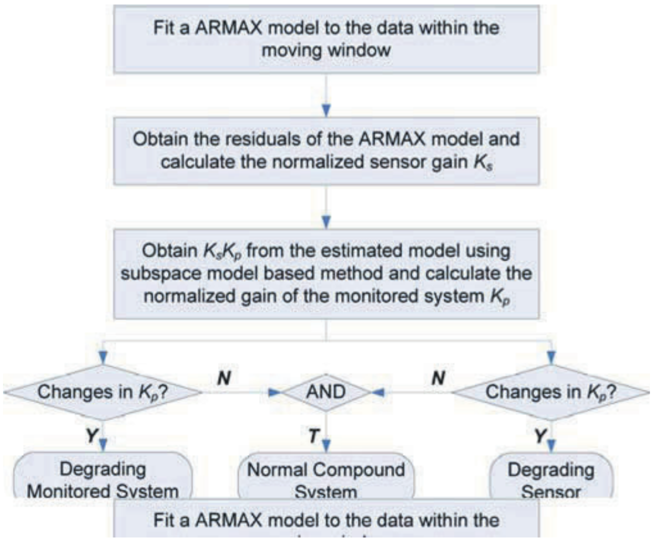


Figure 3. Flow chart of the procedures to detect degraded sensor with gain changes

3.3 Detecting degradation in sensor gain

Due to assumptions of Gaussianity of process $w_p(t)$ and $w_n(t)$, the changes in the identified time constants and gains of the sensor could be statistically interpreted, with severity of those changes being expressed through the overlap of distributions describing behaviour of parameters of the sensor during its normal and degraded operations. Following [12], this overlap between distributions of sensor dynamic parameters under normal and degraded conditions will be referred to as the sensor performance confidence value, ranging between zero and one, with higher CVs signifying a high overlap, and hence performance closer to normal.

4 SIMULATION SCENARIOS – ELECTRIC THROTTLE SYSTEM

To validate the methodology for sensor performance assessment, an electrical throttle system in the simulation environment is introduced. The throttle system is a primary controller that controls the power output in an automobile by controlling the airflow and fuel input to the engine that is actually constrained by the opening and closing of the throttle valve.

Figure 4 shows the simulation model for the compound system with the nominal values of their parameters, which includes an electric throttle as the monitored system and a sensor measuring the throttle angle. In the simulation, the electric throttle is simplified as a 2nd order system, while the angle sensor is modeled as a 1st system with a much smaller time constant than that of the throttle system. Two simulation scenarios for the gradual degradation of the compound system are exampled. 1) The gain of the monitored system and the sensor drop gradually by 40% and 40% simultaneously, respectively. 2) The gain and time constant of the sensor gradually and simultaneously drop by 40% and 40%, respectively.

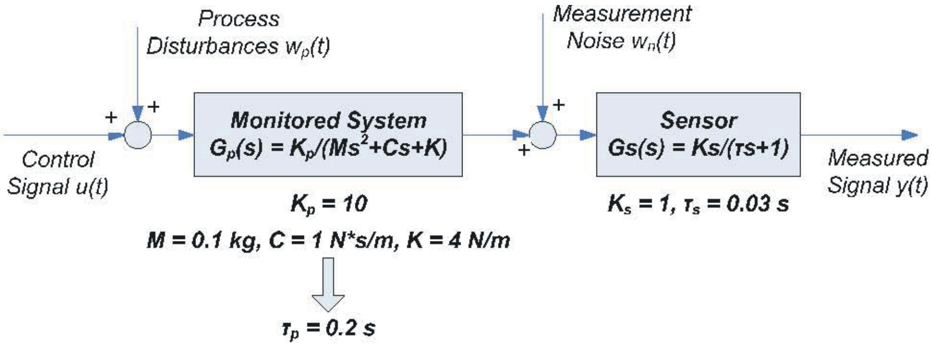
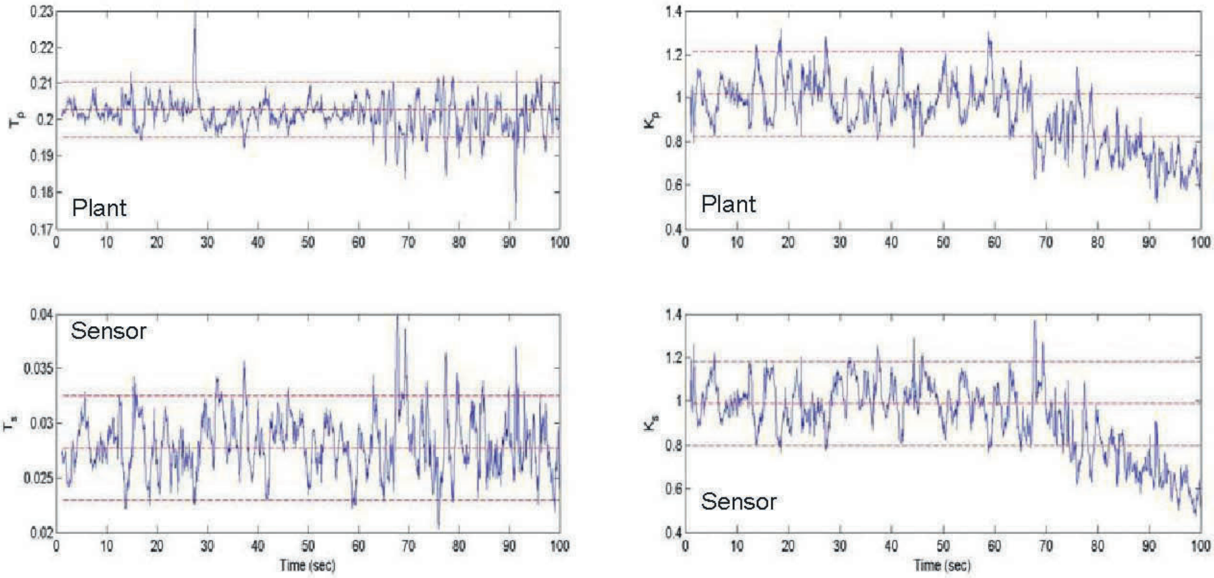
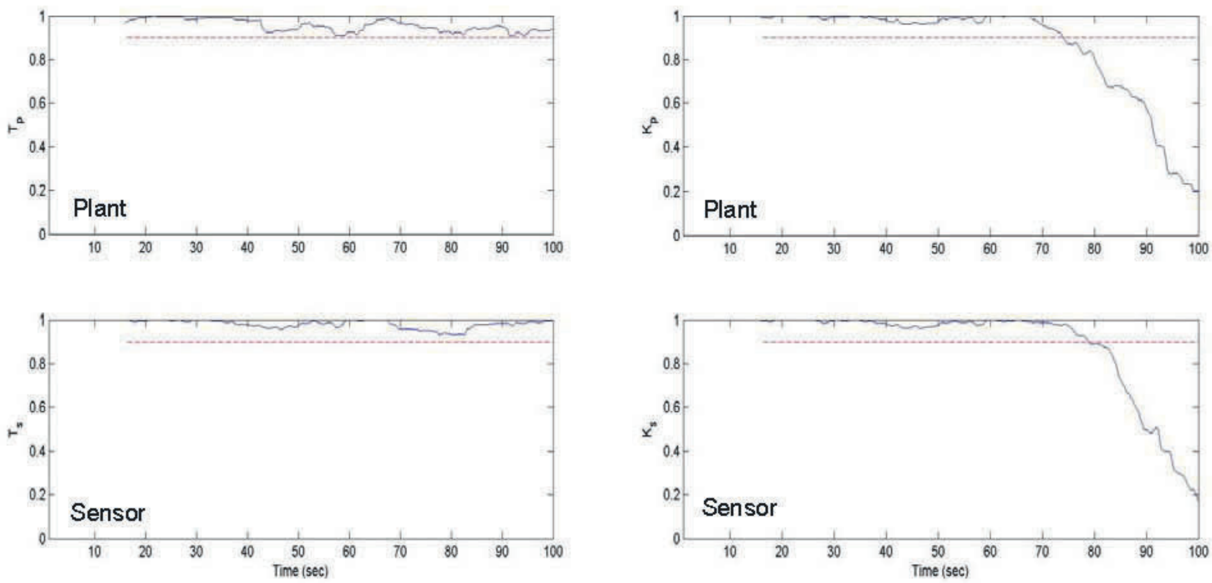


Figure 4. Simulation model and parameters



(a) Estimated Time Constants (b) Normalized Gain Figure 5. Time constants and normalized gains of the plant and the sensor in simulation scenario 1

* *T_p – time constant of the monitored system; T_s – time constant of the sensor;*
* *K_p – normalized gain of the monitored system; K_s – normalized gain of the sensor;*



(a) Estimated Time Constants (b) Normalized Gains

Figure 6. Confidence values of the estimated time constants and normalized gains

As shown in Figure 5(a), the two series of time constants estimated from the data set have mean values at 0.22 *sec* and 0.028 *sec*. According to our assumption that the time constant of the monitored system should be much larger than that of the sensor, we can consider the estimated time constants with mean value of 0.22 *sec* as the time constants of the monitored system, and assigned the other to the sensor. Thus, when degradation occurs in the gains of the throttle system and the angle sensor, the estimated time constants of the monitored system and the sensor oscillate around their mean values without any obvious increasing or decreasing trends. On the other hand, the normalized gains of the monitored system and the sensor, illustrated in Figure 5(b), both show decreasing trends at the end. As shown in Figure 6, the performance CVs associated with the normalized gains drop significantly, while those associated with the time constants indicate normal performance. By setting an ad hoc threshold value as 0.9 in this paper, the performance CVs of the normalized gains of the throttle system and the sensor break through the threshold value, while those of their time constants stay above the threshold.

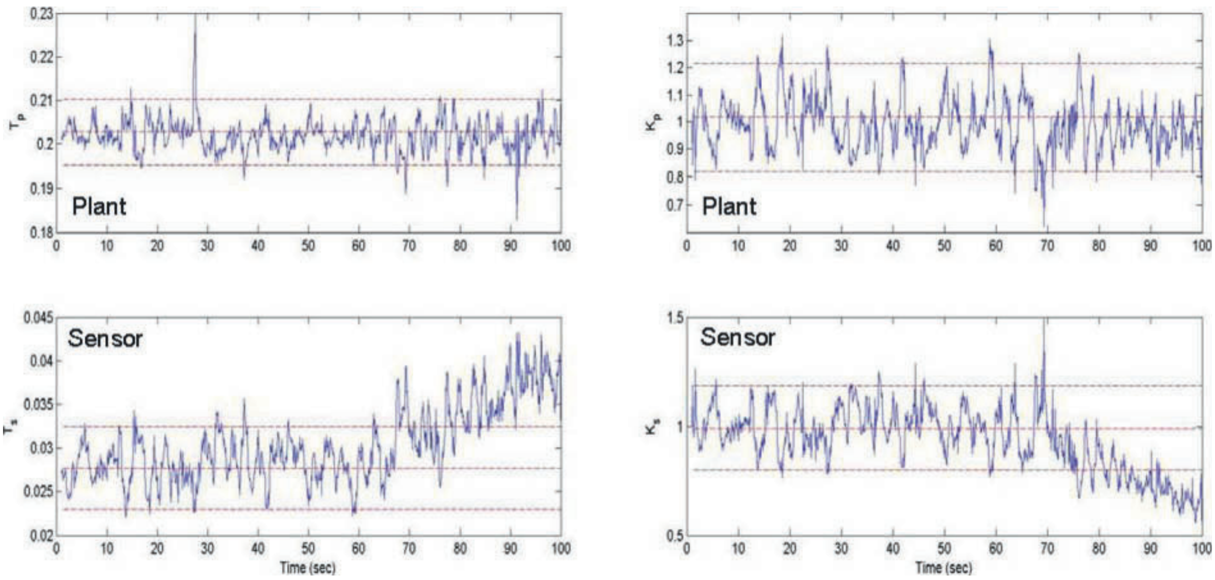
Limits beyond which changes of dynamic sensor parameters are considered significantly can be set using standard Statistical Process Control (SPC) techniques applied directly to the extracted features including the time constants and normalized gains [13]. In this paper, *X* chart is applied with their upper and lower control limits as

$$ULC_x = \overline{X} + 2\sigma_x \qquad LCL_x = \overline{X} - 2\sigma_x \tag{4.1}$$

As shown in Figure 5(a), majority of the estimated time constants of the sensor and the monitored system stay between the upper and lower limits, thus indicating that the dynamic properties of the compound system are under normal conditions. On the other hand, the normalized gain of the sensor and the monitored system, as illustrated in Figure 5(b), consistently break out the low control limit after 75 *sec*, which further verify the diagnostic conclusion that the gains of the compound system are under degradation.

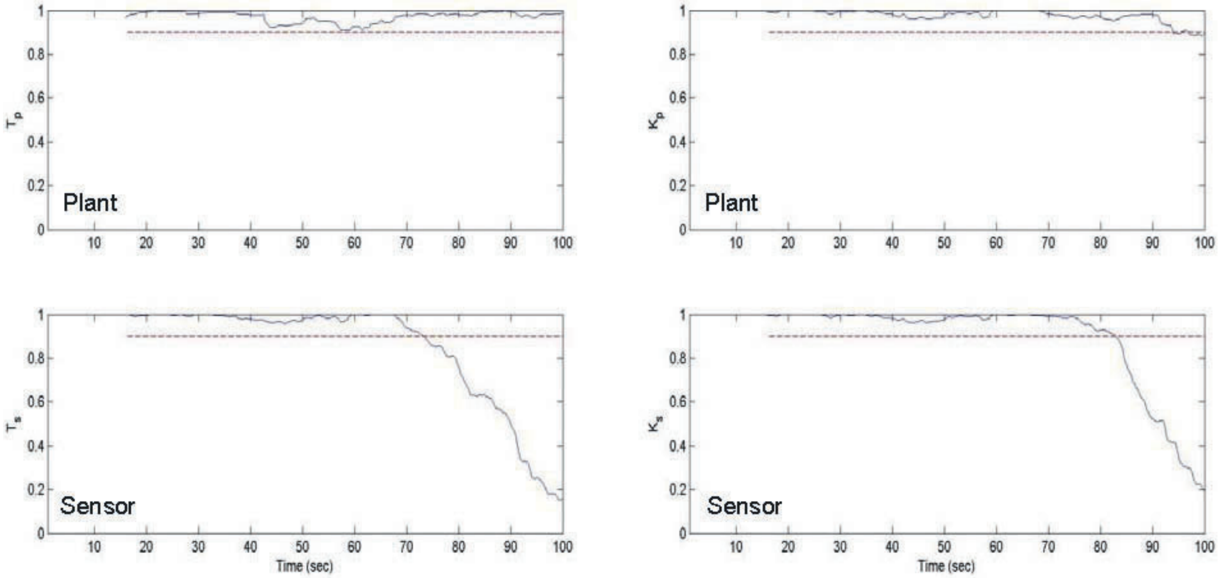
When the gain and time constant of the sensor were changed, the time constants and normalized gains of the throttle mechanism oscillate around their mean values in Figure 7. On the other hand, the time constants of the angle sensor increase and its normalized gains decrease at the end. As shown in Figure 8, the performance CVs associated with the angle sensor drop significantly, while those associated with the throttle system indicate normal performance. Consequently, the performance CVs associated with the angle sensor break through the threshold value, while those associated with the throttle system stay above it.

In addition, as *X* chart is applied to their estimated time constants and normalized gains in Figure 7, the estimated time constants and normalized gains of the sensor consistently break through the control limits after 80 *sec*, which also indicates the degradation of the sensor and verified the conclusion derived from their associated confidence values plots in Figure 8. Thus, the method introduced in this paper is able to detect the degrading sensor and distinguish sensor degradation from that of the plant.



(a) Estimated Time Constants **(b) Normalized Gains**

Figure 7. Time constants and normalized gains of the plant and the sensor in simulation scenario 2



(a) Estimated Time Constants **(b) Normalized Gains**

Figure 8. Confidence values of the estimated time constants and normalized gains

The area under the Receiver Operating Characteristic (ROC) curve, often called AUC, is used to evaluate the discrimination ability of the mainly proposed algorithm for sensor performance assessment. Since the ROC can be represented by plotting the fraction of true positives (TP) vs. the fraction of false positives (FP) for a binary classifier as its discrimination threshold is varied, AUC is generally taken as a summary statistic to evaluate detection accuracy. The AUC of a perfect detector is one because it is able to achieve 100% TP rate and 0% FP rate. On the other hand, the closer the value of AUC comes to 0.5, the less accurate the detector is.

The AUC values in Table.1 are calculated under different degradation modes with step changes in the time constants and gains of the monitored system and the sensor at the levels of 5%, 10%, 15%, and 20%. Since the nominal value of the normalized gain should be one, the normal behaviour of the normalized gains of the monitored system and the sensor are generated according to a Gaussian distribution with mean value of one and standard deviation of the parameters estimated under normal operating conditions. As shown in Table.1, the AUC values for the different degradation modes increase as the degradation level increases, which indicates the intuitively plausible fact that the detection accuracy of the developed algorithm improves as the performance of the monitored system or the sensor deteriorates. In addition, the developed algorithm is able to achieve higher accuracy in detecting changes in the dynamics.

Table 1 : AUC values calculated for four single degradation modes

Degradation Level	*Gain Change*		*Dynamics Change*	
	Monitored System	Sensor	Monitored System	Sensor
5%	0.86778	0.61637	0.99187	1.0000
10%	0.96587	0.85823	0.99996	1.0000
15%	0.99363	0.97445	1.00000	1.0000
20%	0.99839	0.99959	1.00000	1.0000

5 CONCLUSIONS

The method introduced in this paper is able to identify the dynamics of the compound system consisting of a sensor and a monitored plant, and separate the sensor dynamics from that of the monitored plant. As a result, the method is capable of detecting and quantifying senor performance degradation in the compound system without the use of redundant sensing equipment, where either the plants or the sensors monitoring those plants could undergo degradation of their dynamic properties. In addition, the method accomplishes identifying sensor and plant dynamics using inputs observed during normal system operations rather than using special inputs. Consequently, such method is capable of assessing sensor health condition as the system operates, rather than off-line.

6 REFERENCES

[1] D. Djurdjanovic, J. Ni and J. Lee (2002). Time-frequency based sensor fusion in the assessment and monitoring of machine performance degradation. *Proc. of 2002 ASME International Mechanical Engineering Congress and Exposition (ASME IMECE 2002)*, New Orleans, LA.

[2] P. Pike and K. Pennycook (1992). Commissioning of BEMS-A Code of Practice. AH 2/92, BSRIA

[3] P. H. Ibarguengoytia, L. E. Sucar, and S. Vadera (2001). Real Time Intelligent Sensor Validation, *IEEE Transaction on Power Systems* , Vol. 16, No. 4.

[4] A. Badri, F. Gu and A. Ball, Fault-Tolerant Sensor Systems- Alternative Approaches and the Development of a Complex Practical Demonstrator.

[5] S. Yang, J. Yuan, and J. Ni (1996). The Improvement of Thermal Error Modeling and Compensation on Machine Tools by CMAC Neural Network". *Int. J. Mach. Tools Manufact.*, Vol. 36, No. 4, pp. 527-537

[6] G. G. Yen and W. Feng (2001). Winner tak all experts network for sensor validation, *ISA Transactions* 40, pp. 99-110

[7] M. R. Napolitano, D. A. Windon, II, J. L. Casanova, M. Innocenti, and G. Silvestri (1998). Kalman Filters and Neural-Network Schemes for Sensor Validation in Flight Control Systems. *IEEE Transactions on Control Systems Technology*, Vol. 6, No. 5

[8] S. Wang and Y. Chen (2004). Sensor validation and reconstruction for building central chilling systems based on principal component analysis. *Energy Conversion and Management,* 45, pp. 673-695.

[9] G. Kerschen, P. De Boe, J. Golinval and K. Worden (2005). Sensor validation using principal component analysis. *Smart Material and Structure*, 14, pp. 36-42.

[10] Lennart Ljung (1999). System Identification: Theory for the User, Prentice-Hall

[11] S. M. Wu and S. M. Pandit (1993). Time Series and System Analysis with Applications, Krieger Publish Co.

[12] D. Djurdjanovic, J. Lee and J. Ni (2003). Watchdog Agent – An infotronics based prognostics approach for product performance assessment and prediction. *International Journal of Advanced Engineering Informatics, Special Issue on Intelligent Maintenance Systems*, Vol. 17, No. 3 -4, pp. 109-125.

[13] Mick Norton (2005). A Quick Course in Statistical Process Control, Prentice-Hall