

Frontiers  
in  
Artificial  
Intelligence  
and  
Applications

# **MODERN MANAGEMENT BASED ON BIG DATA II AND MACHINE LEARNING AND INTELLIGENT SYSTEMS III**

**Proceedings of MMBD 2021 and MLIS 2021**

Edited by  
Antonio J. Tallón-Ballesteros



**IOS Press**

## **MODERN MANAGEMENT BASED ON BIG DATA II AND MACHINE LEARNING AND INTELLIGENT SYSTEMS III**

It is data that guides the path of applications, and Big Data technologies are enabling new paths which can deal with information in a reasonable time to arrive at an approximate solution, rather than a more exact result in an unacceptably long time. This can be particularly important when dealing with an urgent issue such as that of the COVID-19 pandemic.

This book presents the proceedings of two conferences: MMBD 2021 and MLIS 2021. The MMBD conference deals with two main subjects; those of Big Data and Modern Management. The MLIS conference aims to provide a platform for knowledge exchange of the most recent scientific and technological advances in the field of machine learning and intelligent systems. Both conferences were originally scheduled to be held from 8-11 November 2021, in Quanzhou, China and Xiamen, China respectively. Both conferences were ultimately held fully online on the same dates, hosted by Huaqiao University in Quanzhou and Xiamen respectively.

The book is in two parts, and contains a total of 78 papers (54 from MMBD2021 and 24 from MLIS2021) selected after rigorous review from a total of some 300 submissions. The reviewers bore in mind the breadth and depth of the research topics that fall within the scope of MMBD and MLIS, and selected the 78 most promising and FAIA mainstream-relevant contributions for inclusion in this two-part volume. All the papers present original ideas or results of general significance supported by clear reasoning, compelling evidence and rigorous methods.



ISBN 978-1-64368-224-2 (print)  
ISBN 978-1-64368-225-9 (online)  
ISSN 0922-6389 (print)  
ISSN 1879-8314 (online)

**MODERN MANAGEMENT BASED ON BIG DATA II  
AND MACHINE LEARNING AND INTELLIGENT  
SYSTEMS III**

# Frontiers in Artificial Intelligence and Applications

The book series Frontiers in Artificial Intelligence and Applications (FAIA) covers all aspects of theoretical and applied Artificial Intelligence research in the form of monographs, selected doctoral dissertations, handbooks and proceedings volumes. The FAIA series contains several sub-series, including ‘Information Modelling and Knowledge Bases’ and ‘Knowledge-Based Intelligent Engineering Systems’. It also includes the biennial European Conference on Artificial Intelligence (ECAI) proceedings volumes, and other EurAI (European Association for Artificial Intelligence, formerly ECCAI) sponsored publications. The series has become a highly visible platform for the publication and dissemination of original research in this field. Volumes are selected for inclusion by an international editorial board of well-known scholars in the field of AI. All contributions to the volumes in the series have been peer reviewed.

The FAIA series is indexed in ACM Digital Library; DBLP; EI Compendex; Google Scholar; Scopus; Web of Science: Conference Proceedings Citation Index – Science (CPCI-S) and Book Citation Index – Science (BKCI-S); Zentralblatt MATH.

## Series Editors:

Joost Breuker, Nicola Guarino, Pascal Hitzler, Joost N. Kok, Jiming Liu,  
Ramon López de Mántaras, Riichiro Mizoguchi, Mark Musen, Sankar K. Pal,  
Ning Zhong

## Volume 341

*Recently published in this series*

- Vol. 340. A.J. Tallón-Ballesteros (Ed.), Fuzzy Systems and Data Mining VII – Proceedings of FSDM 2021
- Vol. 339. M. Villaret, T. Alsinet, C. Fernández and A. Valls (Eds.), Artificial Intelligence Research and Development – Proceedings of the 23rd International Conference of the Catalan Association for Artificial Intelligence
- Vol. 338. C. Frasson, K. Kabassi and A. Voulodimos (Eds.), Novelties in Intelligent Digital Systems – Proceedings of the 1st International Conference (NIDS 2021), Athens, Greece, September 30 – October 1, 2021
- Vol. 337. H. Fujita and H. Perez-Meana (Eds.), New Trends in Intelligent Software Methodologies, Tools and Techniques – Proceedings of the 20th International Conference on New Trends in Intelligent Software Methodologies, Tools and Techniques (SoMeT\_21)
- Vol. 336. A. Biere, M. Heule, H. van Maaren and T. Walsh (Eds.), Handbook of Satisfiability – Second Edition
- Vol. 335. M. Nørskov, J. Seibt and O.S. Quick (Eds.), Culturally Sustainable Social Robotics – Proceedings of Robophilosophy 2020 – August 18–21, 2020, Aarhus University and online

ISSN 0922-6389 (print)  
ISSN 1879-8314 (online)

# Modern Management based on Big Data II and Machine Learning and Intelligent Systems III

Proceedings of MMBD 2021 and MLIS 2021

Edited by

Antonio J. Tallón-Ballesteros

*University of Huelva, Spain*



**IOS Press**

Amsterdam Berlin Washington, DC

© 2021 The authors and IOS Press.

This book is published online with Open Access and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).

ISBN 978-1-64368-224-2 (print)

ISBN 978-1-64368-225-9 (online)

Library of Congress Control Number: 2021948323

doi: 10.3233/FAIA341

*Publisher*

IOS Press BV

Nieuwe Hemweg 6B

1013 BG Amsterdam

Netherlands

fax: +31 20 687 0019

e-mail: [order@iospress.nl](mailto:order@iospress.nl)

*For book sales in the USA and Canada:*

IOS Press, Inc.

6751 Tepper Drive

Clifton, VA 20124

USA

Tel.: +1 703 830 6300

Fax: +1 703 830 2300

[sales@iospress.com](mailto:sales@iospress.com)

**LEGAL NOTICE**

The publisher is not responsible for the use which might be made of the following information.

**PRINTED IN THE NETHERLANDS**

# Preface

Data guides the path of applications. Big data technologies are enabling new paths which can deal with information in a reasonable time and arrive at an approximate solution, rather than a more exact result in an unacceptably long time. This is particularly important today as we continue to work against the COVID-19 pandemic.

The conference series Modern Management based on Big Data (MMBD) has held its second edition, whereas the conference series Machine Learning and Intelligent Systems (MLIS) has held its third edition.

MMBD deals with two main branches; those of Big Data and Modern Management. The former includes: data capture and storage; search, sharing, and analytics; big data search, mining and visualization; big data technologies; data visualization; architectures for massive parallel processing; data mining tools and techniques; machine learning algorithms for big data; cloud computing platforms; distributed file systems and databases; scalable storage systems; and big data for business, government and society, among other things. The latter encompasses: modern management; management strategy; management decision making; manufacturing systems; logistic systems; facilities planning; cost analysis; engineering economy; total quality management; management information systems; human factor engineering; and human resources.

The MLIS conference is an annual conference that aims to provide a platform for knowledge exchange of the most recent scientific and technological advances in the field of machine learning and intelligent systems. It also aims to strengthen links in related fields within the scientific community.

This book, which is in two parts, contains accepted papers presented at MMBD 2021 and MLIS 2021. The MMBD conference was originally scheduled to be held from 8–11 November 2021 in Quanzhou, China and MLIS was originally scheduled to be hosted in Xiamen, China on the same dates. The most popular topics in this book concern data analysis and big data applications.

All papers have been conscientiously reviewed by programme committee members, who bore in mind the breadth and depth of the research topics that fall within the scope of MMBD and MLIS respectively. From the some 300 submissions received for MMBD and MLIS, the 78 most promising and FAIA mainstream-relevant contributions have been included in this two-part volume. These present original ideas or results of general significance supported by clear reasoning, compelling evidence and rigorous methods.

I would like to thank all the keynote and invited speakers, as well as the authors and anonymous reviewers, for their effort in making both MMBD and MLIS conferences of the highest standard. We would also like to take this opportunity to express our gratitude to all those people, especially the programme committee members and reviewers, who devoted time to assessing the papers. It is an honour to have been involved from the outset with the publication of these proceedings, which will form part of the prestigious series Frontiers in Artificial Intelligence and Applications (FAIA) from IOS Press. Our particular thanks also go to J. Breuker,

N. Guarino, J.N. Kok, R. López de Mántaras, J. Liu, R. Mizoguchi, M. Musen, S.K. Pal and N. Zhong, the FAIA series editors, for supporting this conference.

Last but not least, any inconvenience caused due to the format change from face to face to virtual is sincerely regretted. Hopefully we will meet face to face at the MMBD2022 and MLIS 2022 conferences next year, the venues for which have not been decided at the time of publication.

September 2021

Antonio J. Tallón-Ballesteros  
University of Huelva (Spain)  
Huelva city, Spain

# Contents

Preface <i>Antonio J. Tallón-Ballesteros</i>	v
<b>Part 1. Modern Management based on Big Data II</b>	
Retailers, All Omni-Shoppers Are Not the Same <i>Nuria Viejo-Fernández</i>	3
The Impact of Tacit Knowledge Acquisition on Firm Performance: The Moderating Effect of Redundant Resources <i>Jing Liu and Fei Yu</i>	16
Opinion-Aware Retrieval Models Based on Sentiment and Intensity of Lexical Features <i>Mohammad Bahrani and Thomas Roelleke</i>	24
Financial Flexibility During the Pre- and Post-Global Financial Crisis Periods <i>Weihua Cui</i>	32
Application of Deep Reinforcement Learning Algorithm in Smart Finance <i>Chunhui Chen and Yichun Zhou</i>	40
Tax Risk Prediction of Real Estate Based on Convolutional Neural Network <i>Meilin Yin and Ning Luo</i>	49
Krylov Subspace Methods for Big Data Analysis of Large Computational Electromagnetics Applications <i>Bruno Carpentieri</i>	57
Industrial Economic Cooperation Between China and Nordic Countries Under the Double Circulation Pattern: Basis and Prospect <i>Nianlin Zhou, Manyuan Jiang and Weixiao Lai</i>	66
Tariff Relief and Promoting Free Trade Agreement Between China and Norway <i>Nianlin Zhou, Jin Guo and Manyuan Jiang</i>	77
The SAPEVO-M-NC Method <i>Sérgio Mitihiro do Nascimento Maêda, Marcio Pereira Basílio, Igor Pinheiro de Araújo Costa, Miguel Ângelo Lellis Moreira, Marcos dos Santos and Carlos Francisco Simões Gomes</i>	89
The Support System for Anomaly Detection with Application in Mainframe Management Process <i>Dominik Strzałka, Alicja Gerka, Bartosz Kowal, Paweł Kuraś, Grzegorz Leopold, Michał Lewicz and Dawid Jaworski</i>	96

Ordering of Warships for the Brazilian Navy Using the New Method: AHP-Gaussian with Pearson's Correlation	104
<i>Carlos Francisco Simões Gomes, Marcus Vinícius Gonçalves Rodrigues, Igor Pinheiro de Araújo Costa and Marcos dos Santos</i>	
Strategic Analysis for the Installation of Field Hospitals for COVID-19 Control: An Approach Based on P-Median Model	112
<i>Leandro de Mattos Bento Soares, Miguel Ângelo Lellis Moreira, Marcio Pereira Basílio, Carlos Francisco Simões Gomes, Marcos dos Santos and Igor Pinheiro de Araújo Costa</i>	
Bibliometric Studies on Multi-Criteria Decision Analysis (MCDA) Applied in Personnel Selection	119
<i>Igor Pinheiro de Araújo Costa, Marcio Pereira Basílio, Sérgio Mitihiro do Nascimento Maêda, Marcus Vinícius Gonçalves Rodrigues, Miguel Ângelo Lellis Moreira, Carlos Francisco Simões Gomes and Marcos dos Santos</i>	
Multicriteria Analysis in Additive Manufacturing: An ELECTRE-MOr Based Approach	126
<i>Paula Drumond, Marcio Pereira Basílio, Igor Pinheiro de Araújo Costa, Daniel Augusto de Moura Pereira, Carlos Francisco Simões Gomes and Marcos dos Santos</i>	
Blockchain Digital Test Certificates for COVID-19	133
<i>Ioannis Karamitsos and Maria Papadaki</i>	
Assessment of National Cybersecurity Capacity for Countries in a Transitional Phase: The Spring Land Case Study	144
<i>Mohamed Altaher Ben Naseir, Huseyin Dogan and Edward Apeh</i>	
Algorithm Selection for Machine Learning Classification: An Application of the MELCHIOR Multicriteria Method	154
<i>Igor Pinheiro de Araújo Costa, Marcio Pereira Basílio, Sérgio Mitihiro do Nascimento Maêda, Marcus Vinícius Gonçalves Rodrigues, Miguel Ângelo Lellis Moreira, Carlos Francisco Simões Gomes and Marcos dos Santos</i>	
Investments in Times of Pandemics: An Approach by the SAPEVO-M-NC Method	162
<i>Sérgio Mitihiro do Nascimento Maêda, Marcio Pereira Basílio, Igor Pinheiro de Araújo Costa, Miguel Ângelo Lellis Moreira, Marcos dos Santos, Carlos Francisco Simões Gomes, Isaque David Pereira de Almeida and Arthur Pinheiro de Araújo Costa</i>	
Role of Big Data Analytics in Belt and Road Initiative (BRI): Multivariate Analysis with Gaussian Distribution of Data	169
<i>Valliappan Raju, Wang Juan, Sandeep Shrestha, Arrunkumar Kalathinathan and KK. Ramachandran</i>	
Research on the Impact of Timing Financing on M&A Performance: The Intermediary Effect Based on Financing Efficiency	178
<i>Qianqian Shang</i>	

Data Analytics Based Risk Management for Students' Performance – A Case Study <i>M. Somasundaram, K.A. Mohamed Junaid, D. Sudha and Sabari L. Umamaheswari</i>	188
Comprehending the Pricing Decision of Online Car-Hailing Services in China: Price Regulation Vs Entry Limitation <i>Jie Yang, Daozhi Zhao, Zeyu Wang and Wanying Jiang</i>	195
A Comparative Evidence of Income Levels Reflecting Gen Z's Digital Payments Intention and Usage <i>Khwanjira Ponsree, Taksin Phongpaew and Phaninee Narhutaradhol</i>	205
Traffic Congestion Model in India by Shock Wave Theory <i>Tsutomu Tsuboi</i>	213
Detection for All Zero Coefficient Blocks in HEVC Based on Uniform Quantizer <i>Nana Shan, Henglu Wei, Wei Zhou and Zhemin Duan</i>	221
Studies on Platform-Type Organization Design Based on Leadership Promotion <i>Xin Ma, Shuhui Li, Wei Yang, Wenhao Yan, Aisheng Liu and Yanting Tan</i>	228
Interdisciplinary Framework: A Building Information Modeling Using Structural Equation Analysis in Lean Construction Project Management <i>Dante Silva, Kevin Lawrence De Jesus, Bernard Villaverde, Andrea Isabelle Enciso, Amanda Nicole Mecija and Justin Owen Mendoza</i>	234
Mitigating Construction Deficiencies: An Impact Analysis for Low-Cost Housing Developments Utilizing Artificial Neural Network <i>Ramene U. Lim, Dante L. Silva and Kevin Lawrence M. De Jesus</i>	241
Research on the Impact of Customer Perceived Value of Haidilao Hot Pot on Consumer Behavior <i>Qiong Li and Anlan Li</i>	248
Study on Construction of Hydrological Monitoring System of the Yangtze River and Application of New Technologies on Flood Monitoring <i>Junya Mei, Bo Zhou and Qiong Wu</i>	256
Management Competency Model: Predictive Neural Network Approach for Empirical Components of Construction Project Proficiency <i>Dante L. Silva, Kevin Lawrence M. De Jesus, Bernard S. Villaverde and Edgar M. Adina</i>	266
Hybrid Neural – Based Model for Predicting the Construction Project Performance of High – Rise Building Projects in Metro Manila, Philippines <i>Jucar Fernandez, Dante L. Silva and Kevin Lawrence M. De Jesus</i>	274
Simulation of Deformation Transfer Coefficient of Pipe Bend Buried Based on Shaking Table Test and Goodman Contact Element <i>Delong Huang, XiaoDong Bai, Xiaoli Chen, Guanyu Xu and Aiping Tang</i>	282
Analysis on the Travel Characteristics of Sharing Bikes Connected with Different Types of Public Transportation Stations <i>Xiaohua Yu, Xiaohui Wang and Yanna Zhao</i>	296

Study on Calculation Method of Contribution Rates of Water Pollution of Trans-Boundary Rivers <i>Xianzhong Zeng and Shouning Hao</i>	305
The Impact of Air Pollution on the Development of Inbound Tourism in Chinese Megacities: The Case Study of Beijing, Guangzhou and Chongqing <i>Nianlin Zhou, Yeli Gu and Manyuan Jiang</i>	317
Research on and Application of Tunnel Structure Defects Prediction Using Machine Learning Methods <i>Bo Shi, Hui Su, Xu Du, Bao Jiao and Lin Wang</i>	329
Evaluation of Water Resources Sustainable Utilization in Wuhan Based on the Model of SUDE <i>Lan Dong</i>	341
An Empirical Study on the Relationship Between Energy Consumption and GDP in China <i>Jian Zhang and Xin Zhou</i>	351
Calculation and Analysis of Carbon Emissions for Green Buildings During Operation Period in Beijing <i>Guojian Wang</i>	360
Location Optimization of Hydrogen Refueling Stations in Hydrogen Expressway Based on Hydrogen Supply Chain Cost <i>Dongxing Wang, Zhe Wang, Fenghui Han, Fan Zhao and Yulong Ji</i>	368
Integrated Urban Energy Planning: A Case-Study Using Optimization <i>Jingbo Guo, Côme Bissuel and Francois Courtot</i>	375
Study of Intelligent Search Engine of Energy Industry Based on BERT Semantic Model <i>Jiayang Li, Hao Li, Ni Yan and Ziyun Chen</i>	385
Game Analysis on Profits of Water Diversion Project Supply Chain Under the Overall Loss of Supply Chain in Undeveloped Areas <i>Jinren Luo, Jiayin Li, Qinyao Fu, Yihu Zhao and Shan Liu</i>	393
Developing a Decision Support System for Water Resources Dispatching <i>Yu Feng, Jijun Xu, Weirong Sheng, Jitian Chen and Yang Hong</i>	406
Research on Key Issues of China's Project Water Price <i>Yanchen Zhou, Xiaoxia Tong, Zhengjie Yin and Lisi Xu</i>	416
The Study of Grading Method and Application Performance of Management Systems of Rural Potable Water Treatment Plants <i>Xinkai Qiu, Cheng Lu, Yinying Zhou and Shuyang Chen</i>	427
Application of Analytic Hierarchy Process (AHP) and Simple Additive Weighting (SAW) Methods in Mapping Flood-Prone Areas <i>Yu Chen</i>	435

Study on Cross-Basin Ecological Compensation Mechanism and Compensation Standard <i>Jinren Luo, Yizhi Sun, Yihu Zhao, Qinyao Fu and Jiayin Li</i>	442
Social and Economic Value Evaluation of Water Resource Based on Fuzzy Mathematics and Emergy Theory <i>Xiang Xu, Yongqiang Wang, Kai Li and Junsong Xin</i>	458
Integrated Flood Risk Assessment System Based on the Set Pair Analysis and Variable Fuzzy Sets for Cascade Dams <i>Yu Chen</i>	471
<b>Part 2. Machine Learning and Intelligent Systems III</b>	
Causal Reasoning Model Based on Medical Knowledge Graph for Disease Diagnosis <i>Ze Xu, Huazhen Wang, Xiaocong Liu, Ting He and Jin Gou</i>	481
The Larger the Better: Analysis of a Scalable Spectral Clustering Algorithm with Cosine Similarity <i>Guangliang Chen</i>	488
About Strong Dependence of the Complexity of Analysis of the Random 3-CNF Formulas on the Ratio of Number of Clauses to the Number of Variables <i>Sergey I. Uvarov</i>	496
Intuitive Contrasting Map for Antonym Embeddings <i>Igor Samenko, Alexey Tikhonov and Ivan P. Yamshchikov</i>	502
DYPLODOC: Dynamic Plots for Document Classification <i>Anastasia Malysheva, Alexey Tikhonov and Ivan P. Yamshchikov</i>	511
Chained Digital Signature for the Improved Video Integrity Verification <i>Linju Lawrence and R Shreelekshmi</i>	520
High Payload Qr-Based Data Hiding Using Secured Compressed Watermark in Polar Domain <i>Indrarini Dyah Irawati, Gelar Budiman, Kholidiyah Masykuroh, Zein Hanni Pradana and Arfianto Fahmi</i>	527
Object and Traffic Light Recognition Model Development Using Multi-GPU Architecture for Autonomous Bus <i>Jheanel Estrada, Gil Opina Jr and Anshuman Tripathi</i>	539
FTR-Based Expert System for Power Generation Units <i>Naftaly Menn and Boris Chudnovsky</i>	548
Context-Sensitive Control of Adaptation: Self-Modeling Networks for Human Mental Processes Using Mental Models <i>Raj Bhalwankar, Laila van Ments and Jan Treur</i>	557
Domain Adaptation for Document Image Binarization via Domain Classification <i>Carlos Garrido-Munoz, Adrián Sánchez-Hernández, Francisco J. Castellanos and Jorge Calvo-Zaragoza</i>	569

Control the Information of the Image with Anisotropic Diffusion and Isotropic Diffusion for the Image Classification <i>Hyun-Tae Choi, Nahyun Lee, Jiwon No, Sangil Han, Jaeho Tak, Hwijn Kim, Haegang Lee, Seonghoon Ham and Byung-Woo Hong</i>	583
Managing Study Stress of College Students Through Personality Traits <i>Ding Ding, Xinyue Liu and Haoran Xu</i>	590
An Adaptive Network Model of the Role of the Gut-Brain Axis in Parkinson's Disease <i>Catherine Goetzinger, Korinzia Toniolo and Jan Treur</i>	596
Studies on the Decrease Mechanisms of Typical Complex Networks <i>Yuhu Qiu, Tianyang Lyu, Xizhe Zhang and Ruozhou Wang</i>	605
Application of Particle Swarm Optimization in Formulation Design of Pig Feed <i>Meijuan Zeng and Hao Li</i>	614
A Comprehensive Energy Optimized Operation Strategy Based on Energy Storage Model <i>Bi-yan Ma, Xiao-dan Zhang and Hao-jie Zhou</i>	622
Soil CEC Predicting Model of Tobacco-Planting Fields in Chenzhou, South Hunan Province <i>YS Xiao, ZH Cao, WJ Li, YH Liao, B He, LJ Li, ZH He and Q Zhong</i>	632
Cutting L1-Norm Distance Discriminant Analysis with Sample Reconstruction <i>Guowan Shao, Chunjiang Peng, Wenchu Ou and Kai Duan</i>	645
L1-Norm Distance Discriminant Analysis with Multiple Adaptive Graphs and Sample Reconstruction <i>Guowan Shao, Chunjiang Peng, Wenchu Ou and Kai Duan</i>	655
Dist Frequent Next Neighbours: A Distributed Galois Lattice Algorithm for Frequent Closed Itemsets Extraction <i>Naomie Sandra Noumi Sandji and Djamal Abdoul Nasser Seck</i>	667
Monitoring and Diagnosis System of Downhole Tubing Leakage <i>Yunpeng Yang, Jianchun Fan, Di Liu and Fanfan Ma</i>	676
Unreadable Segment Recognition of Single-Lead ECG Signals Based on XGBoost: Fusion of Shannon Energy Envelope and Empirical Mode Decomposition <i>Hanshuang Xie, Jiayi Yan, Huaiyu Zhu, Qineng Cao, Yamin Liu, Yun Pan and Fan Wu</i>	688
A Study on the Correlation Between Emotion and the Quality of Life of the Elderly Under the Artificial Intelligent Smile Recognition System: Protocol for a Descriptive Study <i>Ruitong Cai, Ze Zhang, Lu Meng, Kang Ge, Wenhao Yin, Han Qi, Hongxin Zhao and Suning Li</i>	695

**Part 3. Modern Management based on Big Data II – Late Submissions**

Spatio-Temporal Analysis of CO <sub>2</sub> Emission Driving Force in Various Provinces in China Using the Extended STIRPAT-GWR Model <i>Yulin Zhang</i>	703
Investigating the Impact of Carbon Subsidy Policy on the Decision-Making of Remanufacturing Supply Chain <i>Yu Zhang, Tianshan Ma and Syed Abdul Rehman Khan</i>	709
Subject Index	715
Author Index	719

This page intentionally left blank

## Part 1.

# Modern Management based on Big Data II

This page intentionally left blank

# Retailers, All Omni-Shoppers Are Not the Same

Nuria Viejo-Fernández<sup>a,1</sup>

<sup>a</sup> University of Oviedo, Spain

**Abstract.** Taking as a basis the elaboration likelihood model (ELM), this paper evaluates how the way in which the customer searches, evaluates and compares information influences the development of omni-channel behaviour, as well as each of its most common practices, webrooming (researching products online but purchasing products in a physical store) and showrooming (visiting physical stores to check out products and then buying them online). The results obtained from a sample of 939 apparel shoppers using the database constructed for the Spanish retail sector by GfK reflect that compared to one-stop shoppers, omni-shoppers (without distinguishing specific typologies) spend more time and effort planning their decision-making. The combination of physical and virtual channels makes it easier for the consumer to be more involved in the shopping experience and to search, compare and evaluate specific information about the product and/or retailer before the final purchase. This more reflective behaviour in which more time is spent on the consumer journey and more information is handled is what ELM defines as the central information processing route. While it is true that the central information processing route predominates, it is concluded when analysing each of the omni-channel behaviours separately that webroomers are more likely to follow this route, analysing in depth all issues related to the product they want to buy. On the other hand, although showrooming behaviour cannot be associated with the same intensity to the central information processing route, nor can it be associated with a less planned customer journey, like that of e-shoppers who focus only on prices and cost savings associated with the purchase. Showroomers use the internet to learn about retailer-related aspects as well as other consumers' opinions of the product before buying the product from the online store. Taking these results into account, managers should keep in mind the idea that webroomers and showroomers are as different as they are the same. Thus, both the internet and the physical store have to serve as both an information point and a shopping channel. Websites need to be usable and simple so that webroomers can get in-depth information about the retailer's portfolio and showroomers can make a purchase in a few quick steps. On the other hand, the physical store will be a touchpoint where omni-shoppers will enjoy unique experiences, highlighting the sales force that will be key for webroomers and showroomers to develop a stronger bond with the firm and not shop at any other competitor retailer.

**Keywords.** omni-channel behaviour, webrooming, showrooming, information processing routes

## 1. Introduction

Retail is currently a combination of offline and online channels. Academics and business practitioners agree [1] that one of the most relevant developments in the retail environment of recent decades is omni-channel retailing (OCR). In this sense, retailers

---

<sup>1</sup> Nuria Viejo-Fernández, Corresponding author: Department of Business Administration, Avenida del Cristo s/n, 33006 Oviedo, Spain; e-mail: [nuriavjf@uniovi.es](mailto:nuriavjf@uniovi.es)

who want to survive in today's competitive markets will need this business model. OCR is defined as a strategy that integrates the online and offline touchpoints to create a seamless shopping experience that breaks down the barriers between virtual and physical stores and increases engagement during the customer journey [2, 3]. Nowadays, the most common omni-channel behaviours are webrooming and showrooming. While webrooming begins with customers' product information-seeking behaviour through the internet followed by verifying the information by touching and feeling the products in the physical store, where the purchase is finalized, showrooming is the opposite practice [4]: this omni-channel behaviour starts in the physical store, where consumers gather information by 'touching and feeling' products, and ends with buying via the online channel [5].

Both of these behaviours pose a challenge for retailers. Therefore, it is necessary to analyse the drivers that explain what webroomers and showroomers are like so that organizations can segment and adapt their marketing strategies accordingly. The OCR literature from a consumer perspective started out very scarce, as most studies focused on analysing the effect of the combination of physical and virtual channels on companies' turnover, as well as other quantitative consequences of this behaviour. The positive reception of information and communication technologies (ICT) by shoppers without them leaving the physical store (but demanding new formats more suited to their current needs) influenced the OCR literature to focus more deeply on the analysis of consumer behaviour. Initially, demographic and sociocultural variables were the most widely used. However, depending on the product or sector analysed, the studies conducted did not obtain the expected results. These inconclusive results and the complexity that consumer journeys have reached [6] mean that the most recent research in OCR, in addition to the demographic and sociocultural profile, includes psychographic variables [7, 6, 8]. Psychographic factors are broad, and within them, psychological factors such as the self-concept, personality, lifestyle, motivations, interests, attitudes, opinions or values of the consumer, as well as the degree of involvement in the purchase, have significant weight. Involvement reflects the importance that the individual attaches to the specific attributes of the product he or she wishes to purchase, taking into account his or her interests and needs [9]. Therefore, models that analyse consumer behaviour can be taken into account to analyse involvement. Specifically, the elaboration likelihood model (ELM) will be used in this paper.

In this context, this research contributes to the OCR literature by focusing on the study of the effect of different forms of information processing. The results obtained will allow us to identify which type of information processing route is more likely to be associated with each of the omni-channel behaviours, webrooming and showrooming, versus one-stop shopping behaviours (pure offline and pure online).

## **2. Literature review**

The ELM is one of the most widely used theories to study consumption habits. This model, developed by Petty and Cacioppo [10], shows that information processing during the customer journey can be explained by two factors: (1) the customer's motivation to search for information, and (2) the customer's ability to conduct an evaluation. Motivation is related to a customer's degree of involvement. Involvement is revealed as one of the most important variables in explaining and predicting consumer

behaviour because of its influence on the processing of data stored in memory, on the external search for pre-purchase information, and on the formation of attitudes [11].

If customers are more involved with the purchase, their degree of motivation to seek and obtain information will be higher and will be accompanied by a greater investment of time and effort. Thus, information and messages will be analysed through the central route—i.e. in a more critical, rational and in-depth way, developing a proactive behaviour, seeking and paying greater attention to product-related information. In this sense, it is necessary that customers have prior knowledge about the topic or that they are able to connect such information with previous experiences [8, 12]. The alternative route proposed by Petty and Cacioppo [10] is the peripheral route. Customers who follow this route make less cognitive effort as they are less motivated and, consequently, develop lower levels of involvement. Likewise, they are characterized by having less capacity to process the information that allows them to make the purchase decision proactively.

Considering the characteristics of the central information processing route, it seems that it can be associated to a greater extent with the development of omnichannel behaviour. The reason is that omni-shoppers use the different touchpoints, physical and virtual, that organizations make available to them to search for information, compare and evaluate different proposals, and form an opinion on the product they want to buy and the retailer from which they want to make that purchase. Therefore, an omni-shopper can be considered a "smart shopper" [8] since, as opposed to a one-stop shopper, whether pure offline or pure online, he/she combines offline and online channels in such a way that information or experiences that cannot be experienced in one channel can be made up for by using the alternative (for example, touching and feeling the products in the physical store). As a result, his/her customer journey is characterized by being more rational, reflective and longer. The proposed hypothesis is the following:

**H<sub>1</sub>:** A customer who follows the central route to information processing is positively associated with developing omni-channel behaviour.

However, this hypothesis can be qualified by analysing the two most recurrent omnichannel behaviours, webrooming and showrooming. According to Viejo-Fernández et al. [13], webroomers and showroomers have a different psychographic profile. This implies that the motivations that guide their consumer journey, as well as their levels of involvement, will be different despite the fact that both use offline and online touchpoints.

Webrooming behaviour seems to involve a more planned shopping behaviour compared to showrooming behaviour. In this sense, the webroomer uses the internet as a product specialist, gathering as much information as possible about different criteria related to the item he/she wants to buy (characteristics about the product, comparisons with other products...). Similarly, the online channel serves to examine the opinions of other customers and helps form an opinion that is not based on personal assessments or organizational information that may be biased, but on real experiences of other individuals. The information gathered on the internet means that the webroomer has a firm and decided idea about the product he/she wants to buy when he/she goes to the physical store. However, the salesroom plays an important role for these omni-shoppers, insofar as it is the touchpoint chosen to make the purchase of the product they wish to buy. According to McKinsey and PwC [14, 15], retailers need to provide a so-called "phygital experience" to engage customers. The phygital experience is about fostering a digital presence as well as the ability to buy online freely in store. Because webroomers

are more involved in shopping and their leitmotiv is to seek as much information as possible to evaluate alternatives and make the right purchase decision, it is important for retailers to have tablets or other electronic devices available. By placing these technological tools as a part of the furniture at strategic points in the salesroom and/or next to the products, webroomers can expand the information as they see the product they want to purchase in the store (for example, prices in other competing retailers, characteristics of other similar products, need for accessories, colours, sizes...). The use of smartphones should also be allowed. Many retailers include frequency jammers in their shops, but far from making people buy, the opposite effect is achieved. In this sense, the smartphone will be used to help webroomers complete the information inside the store after seeing the products or listening to the salesperson [16]. The key is for retailers to be well positioned in the online channel.

Similarly, the salesperson is also key to closing the purchase. Webrooming consumers arrive with a firm idea of what they want to buy and with knowledge that can sometimes exceed the knowledge of salespeople, provided that the sales staff prolong the stay of these omni-shoppers in the physical store. The aim is to persuade them to change their opinion slightly (but not completely). Thus, retailing intelligence tools can help. For example, consumer recognition systems will help retailers to provide a personalized response according to the consumer's profile. To do this, retailers need to have a database of their shoppers in which their purchase history, average spending, etc. are collected. In this way, the salesperson, using a tablet, will be able to know the consumer's preferences and adapt to them, advising them in a way that the online channel cannot. Likewise, the salesperson can make use of commercial techniques such as cross-selling, which consists of offering the buyer a complementary product or accessories that complement the product they want to buy. They can also apply up-selling, which consists of suggesting to the buyer a product that will improve the initially chosen one: although it makes the purchase more expensive, the salesperson can argue that the higher outlay is worth it because it will exceed the customer's initial expectations.

Showroomers have a different psychographic profile to webroomers. While showroomers combine offline and online channels, they are e-shoppers. According to Solomon [17], two types of purchase motivation can be distinguished: utilitarian motivations and hedonistic motivations. Utilitarian motivations occur when consumers are oriented towards seeking functional or economic benefits. Hedonistic motivations are associated with the emotional component of the consumer, and the search for novel and pleasurable experiences during the purchase. If these motivations are applied to the omnichannel environment, the search for economic or functional benefits will be associated with the channel that saves on all the costs associated with shopping: time, effort, travel, price, etc. [18]. These features seem to characterize the online channel. Among the advantages of the internet as a shopping channel, reports analysing e-commerce reflect convenience, comfort or the availability of a varied product assortment at affordable prices. Therefore, since showroomers shop online, their motivations for purchasing the product will be utilitarian; consequently, since motivation and level of involvement are two closely linked psychographic variables, according to Rapp et al. and Yurova et al. [19, 20], showrooming consumers will have a lower level of involvement compared to webroomers. So, showroomers can be more easily persuaded by sales staff than webroomers, emphasizing the most attractive attributes of the brand or the physical store, which will be associated with utilitarian attributes.

Regardless of the previous comments, showroomers are omni-shoppers. Therefore, showrooming consumers cannot be associated with either the peripheral route of information processing or the central route of information processing. They will be halfway between the two routes, for, although utilitarian motivations weigh a little more than hedonistic motivations, hedonistic motivations are latent. In this way, the experience lived in the physical store while informing will be important. Thus, enhancing sensory stimulation through the retailer's promotional activity, novelties, merchandising, or the availability of products in the shop, may positively influence these omni-shoppers to make the purchase in the retailer's virtual store [21, 8], avoiding so-called competitive showrooming [22].

A comparison drawn between webrooming and showrooming leads us to make the following hypothesis:

**H<sub>2</sub>:** Webroomers are more likely to use the central route to information processing than showroomers.

### 3. Methodology

To test the proposed hypotheses, the empirical research used the e-Commerce Observatory for the Spanish market conducted by GfK. Among the different retail sectors analysed in this database, this paper chooses the apparel sector because, in addition to its economic relevance for any developed country, it is also one of the sectors with the highest OCR index [23]. The GfK Spain database provides detailed information on the customer journey based on a sample of 939 customers who have purchased fashion products.

To measure the specific type of behaviour (omni-channel versus one-stop shopping), an endogenous multi-response variable (*typebehaviour*) with three mutually exclusive alternatives is used. Values 1 and 2 define individuals who follow an omni-channel behaviour, the first describing webroomers and the second showroomers. Value 3 includes all individuals who develop one-stop shopping behaviours (pure offline customers and pure online shoppers), this being the alternative chosen as a reference. In this case, the two types of one-stop shopper are not considered separately as the two individual samples are small (together they add up to 87 observations, 42 pure offline shoppers, 45 pure online shoppers). The smaller the sample, the more imprecision there will be in the results (and the wider the confidence intervals of the studied parameters). Thus, the differences will have to be larger in order to reach statistical significance. Thus, even if there is a real difference, its existence cannot be assured because the sample is too small, losing the opportunity to demonstrate differences that, although small, may be relevant.

To evaluate the information processing routes, the central route and the peripheral route proposed in the ELM, the research used three blocks of variables that were part of the GfK Spain database (Table 1). The first two blocks refer to the decision criteria related to the product and the retailer where the purchase is made, respectively. The third block of variables measures the degree to which a customer plans his or her purchase. All the items in these blocks are dichotomous variables. The value 1 refers to when the respondent has answered affirmatively to have carried out the action about which he/she is asked; the value 0 reflects the opposite.

**Table 1.** Measurement of information processing

Block	Code
I. Determinant decision criteria relating to the product. <i>"Regarding the product you purchased, what information motivated you to make the purchase?"</i>	
Price	<i>priceproduct</i>
Characteristics	<i>characteristicsproduct</i>
Comparisons with other products	<i>comparativeproduct</i>
Reviews from customers	<i>womcustomer</i>
II. Determinant decision criteria relating to the retailer. <i>"Regarding the retailer from which you made the purchase, could you state what information motivated you to make the purchase?"</i>	
Product availability	<i>availabilitystore</i>
Price	<i>pricestore</i>
e-Commerce services	<i>onlineservices</i>
III. Purchasing process. <i>"Referring to the purchasing process, say which one of these sentences more closely reflects your usual product purchases."</i>	
I researched thoroughly and knew exactly what I wanted	<i>thickinformation</i>
I spent time visiting shops and on shopping	<i>thickcheck</i>
I went to the store only to buy the product	<i>exclusivepurchase</i>
I had been thinking about buying the product for some time	<i>musepurchase</i>

Source: e-Commerce Observatory, GfK Spain

As proposed in the previous section, it can be assumed that those individuals who have followed the central route to information processing will analyse in great depth all the information related to the product they wish to purchase before making their acquisition. In this sense, they will not only analyse the characteristics, technical specifications or attributes, and the price in detail, but also those decision criteria related to a product that may be of greater relevance in determining their purchase (the possibility of making comparisons or the possibility of knowing the opinions of other consumers). Also, consumers who engage in such a proactive search for information will develop longer shopping journeys, where the buyers spend time and effort visiting different stores, comparing products and the information provided by the sales staff.

On the other hand, consumers who choose to follow the peripheral route to information processing are expected to focus their purchase decision on criteria related to the retailers selling the product and all the services they provide. As opposed to who have followed the central route to information processing the previous one, since they are not oriented to seek in-depth information about the products, their consumer journey will tend to be less prolonged and they will visit stores only with the objective of buying the article that is offered at the best price and/or at the retailer that they consider provides the best service.

#### 4. Results

Table 2 shows the results of the estimated binomial logit model using statistical software STATA 12. The results show a good overall significance of the parameters and give some support to hypotheses  $H_1$  and  $H_2$  proposed in this paper. If one looks at

the three blocks of variables used to measure information processing (central versus the peripheral route), there are a large number of significant results (highlighted in bold). In particular, 90 per cent of the items used have significant results that support the hypotheses posed. In this sense, when omni-channel behaviour is analysed without distinguishing between specific typologies (**H<sub>1</sub>**), it can be seen that omni-shoppers, compared to one-stop shoppers, have a longer decision-making period in which they invest time and effort in searching and evaluating the alternatives linked to the product they wish to acquire, as well as the characteristics linked to the retailer where they wish to make their purchase. Omni-shoppers analyse in depth all the characteristics related to the product they wish to buy, as well as the opinions of other consumers who have already bought the product, which in many cases serve to decide on the purchase (peer-to-peer opinions currently carry more weight than one-way communication from companies). Similarly, omni-shoppers consider it relevant to obtain information on the price or availability of the product they want in retailers' physical and virtual shops. Moreover, in the case of omni-channel consumers who are e-shoppers, it is of vital importance to know what kind of online services firms provide in their e-commerce.

In summary, the results show that omni-shoppers (without distinguishing between specific typologies) follow the central information processing route. Therefore, their level of involvement in the purchase is higher than those consumers who only use a single channel (either offline or online). However, when analysing the two most recurrent omnichannel behaviours, webrooming and showrooming, there are some differences that are worth pointing out. The results in Table 2 show that webroomers are more likely to be associated with the central information processing route than showroomers, as proposed in hypothesis **H<sub>2</sub>**. It seems that consumers who engage in webrooming behaviour inform themselves in depth prior to purchase, investing time in evaluating the different alternatives before making a firm decision and going to the physical store to buy. Thus, when analysing the block of variables relating to the product, we can see how vital it is for webroomers to search for information about the product they want to buy. All items are significant and positive (product price, characteristics, comparisons with other products, and opinions of other consumers). They also look for information on which retail outlets have availability of the goods they want and their prices.

**Table 2.** Binomial logit model estimated for webroomers and showroomers versus one-stop shoppers (pure offline and pure online customers)

Variable	Webrooming		Showrooming	
	Coefficient		Coefficient	
	(Robust Std. Err.)	(Robust Std. Err.)	(Robust Std. Err.)	(Robust Std. Err.)
Block I. Determinant decision criteria relating to the product				
<i>priceproduct</i>	<b>1.278***</b> (0.293)		-0.158 (0.301)	
<i>characteristicsproduct</i>	<b>0.527***</b> (0.252)		-0.300 (0.351)	
<i>comparativeproduct</i>	<b>0.483**</b> (0.350)		0.109 (0.462)	
<i>womcustomer</i>	<b>0.785**</b> (0.361)		<b>0.702*</b> (0.372)	

## Block II. Determinant decision criteria relating to the retailer

<i>availabilitystore</i>	<b>0.382*</b> (0.178)	<b>0.592**</b> (0.387)
<i>pricestore</i>	<b>0.496**</b> (0.201)	<b>0.452**</b> (0.315)
<i>onlineservices</i>	0.901 (0.274)	<b>1.201***</b> (0.555)

## Block III. Purchasing process

<i>thickinformation</i>	<b>0.402**</b> (0.201)	0.342 (0.302)
<i>thickcheck</i>	0.198 (0.200)	0.210 (0.315)
<i>exclusivepurchase</i>	0.063 (0.199)	<b>0.263*</b> (0.352)
<i>musepurchase</i>	<b>0.206*</b> (0.215)	0.122 (0.296)

\*p < 0.10; \*\*p < 0.05; \*\*\*p < 0.01.

Showroomers, on the other hand, do not seem to have as long a purchase journey as webroomers. Therefore, it can be considered that their decision-making process is more impulsive. As shown in Table 2, showrooming consumers seem to visit only the online store of the retailer they are going to buy from. The retailer's information is important for these omni-shoppers, who look at the price the company sets for the product they want to buy, the availability of the merchandise in the company, and the online services the company provides. As far as product-related variables are concerned, the results obtained show that the opinions of other consumers are decisive when it comes to making a purchase, perhaps partly due to the low level of involvement in the search for information.

To complete the results in Table 2, a Chi-squared test is performed to analyse the differences between webroomers and showroomers (Table 3).

**Table 3.** Chi-squared test estimated for webroomers and showroomers versus one-stop shoppers (pure offline and pure online customers)

Variable	Webrooming (436)	Showrooming (416)	One-stop shoppers (87)	Chi-squared
Block I. Determinant decision criteria relating to the product				
<i>priceproduct</i>	315 (72.25%)	322 (77.40%)	71 (81.61%)	19.36 (p=0.00)
<i>characteristicsproduct</i>	408 (93.58%)	372 (89.42%)	66 (75.86%)	17.72 (p=0.00)
<i>comparativeproduct</i>	386 (88.53%)	304 (73.07%)	48 (55.20%)	6.61 (p=0.04)
<i>womcustomer</i>	315 (72.25%)	402 (99.01%)	11 (12.65%)	1.86 (p= 0.07)
Block II. Determinant decision criteria relating to the retailer				

	387	396	48	22.58
<i>availabilitystore</i>	(88.76%)	(95.19%)	(55.17%)	(p=0.00)
<i>pricestore</i>	403	411	13	12.76
<i>onlineservices</i>	(92.43%)	(98.80%)	(14.94%)	(p=0.02)
	306	407	7	20.39
	(70.18%)	(97.83%)	(8.05%)	(p=0.00)
<b>Block III. Purchasing process</b>				
	387	315	9	7.75
<i>thickinformation</i>	(88.76%)	(75.72%)	(10.35%)	(p=0.02)
<i>thickcheck</i>	228	206	6	8.76
<i>exclusivepurchase</i>	(52.30%)	(49.52%)	(6.89%)	(p=0.09)
	301	389	17	5.39
<i>musepurchase</i>	(69.03%)	(93.51%)	(19.54%)	(p=0.07)
	367	360	22	15.37
	(84.17%)	(86.54%)	(25.28%)	(p=0.08)

## 5. Discussion

The objective of this research was to analyse how the way in which the customer searches, evaluates and compares information influences the development of omnichannel behaviour, as well as each of its most common practices, webrooming and showrooming. For this, the ELM was taken as a reference, whose postulate is to analyse the information processing routes, central route versus peripheral route.

This work has theoretical and managerial implications. In terms of theoretical implications, the results of this research support the findings of other similar studies (e.g. [8, 12]) and contribute to the current state of the OCR literature. Omni-shoppers emerge as a category of shoppers who have a longer decision-making process than one-stop shoppers. Combining channels and touchpoints, offline and online, implies a more reflexive behaviour that will be accompanied by a greater investment of time and effort. It is clear that this consumer can be defined as a "smart shopper", as he/she looks for information that can help him/her make the best decision and avoid the risk of making a mistake. Thus, information that cannot be found in the offline channel can be obtained from the online channel. Omni-shoppers will be informed in depth about all aspects related to the product in question, analysing its characteristics, the price range in which it sits, whether there is the possibility of complementing the product with accessories, and whether there are complementary products. They also use the comments and opinions of other buyers as a reference to make their decisions. As an omni-shopper, they will use traditional word of mouth (WOM) and e-WOM as a reference point. Peer-to-peer feedback is nowadays often more influential than advertising or retailer communications. In fact, according to López-López et al. [24], negative experiences are more resistant to change and those who experience them will want to communicate them with greater fervour to their peers as a way of getting back at the company.

Similarly, it is important to find out about the retailer where the product is going to be purchased. Knowing the price set for the product that the consumer wishes to buy and the quantity of product available is relevant, above all, to avoid having to wait, or to wait as little time as possible for the item that he or she wishes to buy. This is especially relevant in the apparel sector, where some models, especially the most popular ones advertised by influencers, tend to sell out quickly. Omni-shoppers have increasingly complex consumer journeys [6], so many of them will have a greater affinity with ICT and will want to buy online. Thus, it is relevant for them to be informed about the services offered by e-commerce retailers. Knowing the delivery policy, shipping costs, or the possibility of making returns online and in the physical shop of the firm is something relevant for e-shoppers who are characterized by the fact that their purchase is guided by functional motivations.

While the above considerations are true, the omni-shopper is not a homogeneous consumer. The ways in which omni-shoppers combine offline and online touchpoints can be very varied: in fact, different nomenclatures have emerged in recent years to describe the most common types. However, academic literature and professional reports determine that the most recurrent omnichannel behaviours are webrooming and showrooming. According to Viejo-Fernández et al. [8, 12, 13], these omni-shoppers differ in their psychographic profiles, especially in the motivations that guide their purchase and, consequently, their involvement in it. The webroomer shows more planning behaviour and invests more time and effort in the search for and evaluation of information, especially related to the product he/she wishes to purchase. Although the webroomer is also informed about some issues related to the retailer (product availability and price), these issues are of vital importance for showroomers. Thus, as webroomers do not go to the salesroom with a definite idea of what they want to buy, being able to see the product, the price offered by the retailer and the services provided (especially online services) can decide the purchase. This implies that decision-making is more impulsive for showrooming customers.

The results obtained in this research help retailers to implement appropriate segmentation strategies to make OCR successful. Defining omni-shoppers as homogeneous consumers is wrong. In fact, the latest work on consumer behaviour and business practitioners is oriented towards analysing purchase journeys in order to try to give the appropriate response to each type of customer. This means that there is no uniformity in the way consumers search for and evaluate information, nor where they buy. The purchase journey may start offline and end online or vice versa, or it may start and end at the same physical (or online) touchpoint, using different virtual (or offline) touchpoints in between. While companies must be prepared for the complexity of the consumer journeys that are beginning to emerge, today they must focus on responding to two types of consumer in particular, webroomers and showroomers.

While webroomers emerge as individuals who develop a longer customer journey over time, looking for in-depth information about the product they want to buy, showroomers appear to be omni-shoppers with less planned purchasing behaviour. They look at the characteristics of the retailer—that is, showrooming consumers choose the company that offers the product at the best price with a certain level of quality. It can be considered that webroomers and showroomers are two sides of the same coin. Therefore, retailers' physical and virtual stores must be prepared to provide a seamless response to each of these two omni-shoppers insofar as they do not know a priori whether they will use the website or the physical store to search for information or to buy.

Therefore, although the internet is linked to elements of a functional nature that are aimed solely at shoppers guided by utilitarian motivations such as saving time and effort, comfort or convenience, today's retailers must also promote elements that serve those omni-shoppers who have a greater involvement in the purchase and hedonistic motivations, such as webroomers. For these omni-channel consumers, the website where they look for product information is the first contact they have with the retailer. Therefore, the company's website should have a simple but eye-catching structure that is easy to navigate and use. It should also be accessible through any static or mobile device, the so-called responsive design. Colours should capture the user's attention without distracting them from what they want to look at. Images are essential so that the product can be seen in full detail. For this, the image needs to be enlarged and visible from different angles. It is also important that the virtual store has quality information about the portfolio of products that the firm sells, details of the price and main characteristics of the item in question, shows complementary products, whether it needs accessories, available stock, etc. The information should not be limited to the product: the services provided by the retailer, such as contact forms, delivery times, advice and personalized recommendations, opinions of experts and influencers, collection of shipping costs, flat rates or free shipping, are of great interest, especially for those consumers who buy online, such as showroomers. Likewise, the purchase process must be carried out in a few simple steps so that e-shoppers do not have a high abandonment rate.

For their part, physical stores must have a more up-to-date concept than they have had so far. This is what PwC [15] calls the phygital experience, which consists of incorporating ICT to achieve a strong bond with the shopper and, consequently, their engagement. According to Mosquera et al. and Viejo-Fernández [16, 25], the inclusion of tools such as tablets next to products, augmented reality, virtual fitting rooms or i-beacons can improve the shopping experience since information on the retailer's offer is available at all times. Retailers are interested in webroomers and showroomers staying as long as possible in the physical store—in the case of webroomers, who come to the salesroom with a firm idea and attitude towards what they want, to redirect them (without totally changing their mind) to buy a product more adapted to their needs which implies a higher outlay; and in the case of showrooming consumers, so that they can contemplate the advantages of buying the product from this firm, avoiding purchase in the online shop of a competing retailer or competitive showrooming [22]. For this reason, in addition to ICT, it is necessary that apparel retailers have a salesperson adapted to omni-shoppers. This is the so-called "salesperson 4.0" [26]. In addition to the traditional skills of a salesperson, such as empathy, listening skills and effective communication, it is necessary to add other skills that facilitate sales in the current environment where the weight of new technologies is so prominent. According to Gil [26], this salesperson must be an expert in digital marketing, given the hyperconnectivity of today's consumers, know the social media in which they find information before buying the product, and know what they are commenting on and what they prefer or reject. Knowledge management is also relevant, as having information about what the consumer thinks and wants allows the company to adapt its offer. The salesperson also needs to be creative, understanding creative as synonymous with resilience, and be able to adapt to the challenges of persuading today's consumers who, thanks to ICT, are increasingly informed about the characteristics of the product they want to buy and its price. Finally, the "talent 4.0" competence implies that the salesperson must seek excellence of results, which involves making sales so that retail

companies meet their commercial objectives through the different physical and virtual channels they make available to consumers, preventing them from turning to the competition.

Finally, for showroomers characterized by a less planned decision-making process synonymous with a more impressionable behaviour, any element that activates their sensory stimulation should be promoted. Consequently, commercial strategies aimed at the showrooming consumer would involve the retailer encouraging the experience in the physical store and facilitating online shopping in store, encouraging salesperson 4.0 to take the initiative and show this omni-shopper in the physical store the products where the retailer is well positioned in terms of price so that showroomers can purchase them if they wish to do so in the virtual channel.

### *5.1. Limitations and future research*

While the study has an extensive database for empirical research, it has two important limitations. On the one hand, the e-Commerce Observatory is a cross-sectional study that focuses on a specific time period. On the other hand, despite the fact that the database was created by an expert company, GfK Spain, not all the variables it incorporates are in line with what it is intended to measure.

It should be noted that the research has focused on a specific sector, apparel retailing. Also, the study has basically focused on the two most common types of omni-channel behaviour, webrooming and showrooming. However, it is possible to identify more specific behaviours, such as click and collect or drive-through, that have not been contemplated in this study. Future research could aim to consider these omni-channel behaviours. Understanding which specific touchpoints, offline and online, the customers use at each stage of the purchase, why they make use of those touchpoints and not others, and how they behave at each stage would further advance the OCR literature.

It would also be of great value to study the factors that condition omnichannel behaviour, especially those related to psychographic variables, as well as the consequences from a cognitive-affective perspective, since, although there are studies that analyse these variables, they are still scarce and incipient, both separately and jointly. Finally, future research could include other retail sectors in order to verify whether the results follow a pattern. In this sense, it would be particularly interesting to consider omnichannel behaviour in the case of sectors where OCR has a low index, such as the toy sector.

## References

- [1] World Retail Congress (2020). Omni-Channel Retailing: The new normal. Available at: <https://www.worldretailcongress.com/partner-content/omni-channel-retailing-the-new-normal>.
- [2] Grewal, D., Motyka, S., Levy, M. The Evolution and Future of Retailing and Retailing Education. Journal of Marketing Education. 2018; 40(1): 85-93.
- [3] Flavián, C., Gurrea, R., Orús, C. Combining channels to make smart purchases: The role of webrooming and showrooming. Journal of Retailing and Consumer Services. 2020; 52: 101923.
- [4] Aw, E.C.X., Basha, N.K., Ng, S.I., Ho, J.A. Searching online and buying offline: Understanding the role of channel-, consumer-, and product-related factors in determining webrooming intention. Journal of Retailing and Consumer Services. 2021; 58: 102328.
- [5] Fiestas, J.C., Tuzovic, S. Mobile-assisted showroomers: Understanding their purchase journey and personalities. Journal of Retailing and Consumer Services, 2021; 58: 102280.

- [6] Radzevičė, J., Banytė, J. Driving Factors of consumer irrationality in Omnichannel Consumer Behaviour. *Marketing and Smart Technologies*, Singapore: Springer; 2020.
- [7] Nakano, S., Kondo, F.N. Customer segmentation with purchase channels and media touchpoints using single source panel data. *Journal of Retailing and Consumer Services*. 2018; 41: 142-152.
- [8] Viejo-Fernández, N., Sanzo-Pérez, M.J., Vázquez-Casielles, R. Webroomers versus showroomers: Are they the same? *Journal of Business Research*. 2018; 92: 300-320.
- [9] Drichoutis, A.C., Lazaridis, P., Nayga, R.M. An assessment of product class involvement in food purchasing behavior. *European Journal of Marketing*. 2007; 41(7/8): 888-914.
- [10] Petty, R.E., Cacioppo, J.T. Communication and persuasion: Central and peripheral routes to attitude change. New York: Springer-Verlag; 1986.
- [11] Cabafíero, C.P. La implicación del consumidor en la adquisición de bienes y servicios. *EsicMarket*. 2006; 123: 93-113.
- [12] Viejo-Fernández, N., Sanzo-Pérez, M.J., Vázquez-Casielles, R. Different kinds of research shoppers, different cognitive-affective consequences. *Spanish Journal of Marketing-ESIC*. 2019; 23(1): 45-68.
- [13] Viejo-Fernández, N., & Sanzo-Pérez, M. J. Combining internal and external drivers to understand omni-channel behavior. *International Journal of Business Marketing and Management*. 2020; 5(11).
- [14] McKinsey (2019). The State of Fashion 2020. Available at: <https://www.mckinsey.com/industries/retail/our-insights/the-state-of-fashion-2020-navigating-uncertainty>.
- [15] PwC (2018). Rethinking retail: The role of the physical store. Available at: <https://www.pwc.be/en/documents/20180627-rethinking-retail.pdf>.
- [16] Mosquera, A., Olarte-Pascual, C., Juaneda-Ayensa, E., Sierra Murillo, Y. The role of technology in an omnichannel physical store: Assessing the moderating effect of gender. *Spanish Journal of Marketing*. 2018; 22(1), 63-82.
- [17] Solomon, M.R. Comportamiento del consumidor, New Jersey: Prentice Hall; 1997.
- [18] Pookulangara, S., Hawley, J., Xiao, G. Explaining multichannel consumer's channel-migration intention using theory of reasoned action. *International Journal of Retail & Distribution Management*. 2011; 39(3), 183-202.
- [19] Rapp, A., Bakera, T.L., Bachrachb, D.G., Ogilviea, J., Beitelspacherc, L.S. Perceived customer showrooming behavior and the effect on retail salesperson self-efficacy and performance. *Journal of Retailing*. 2015; 91(2): 358-369.
- [20] Yurova, Y., Rippé, C.B., Weisfeld-Spolter, S., Sussan, F., Arndt, A. Not all adaptive selling to omniconsumers is influential: The moderating effect of product type. *Journal of Retailing and Consumer Services*. 2016; 34, 271-277.
- [21] Bezes, C. Identifying central and peripheral dimensions of store and website image: Applying the elaboration likelihood model to multichannel retailing. *Journal of Applied Business Research*. 2015; 31(4): 14-53.
- [22] Gensler, S., Neslin, S.A., Verhoef, P. C. The showrooming phenomenon: it's more than just about price. *Journal of Interactive Marketing*. 2017; 38: 29-43.
- [23] FashionUnited (2020). Global fashion industry statistics-International apparel. Available at: <https://fashionunited.com/global-fashion-industry-statistics/>.
- [24] López-López, I., Salvador Ruiz-de-Maya, S. Warlop, L. When sharing consumption emotions with strangers is more satisfying than sharing them with friends. *Journal of Service Research*. 2014; 17(4): 475-488.
- [25] Viejo-Fernández, N., Sanzo-Pérez, M.J., Vázquez-Casielles, R. Is showrooming really so terrible? Start understanding showroomers. *Journal of Retailing and Consumer Services*. 2020; 54: 102048.
- [26] Gil, J. (2017). Competencias clave del Vendedor 4.0. Available at: <https://es.linkedin.com/pulse/competencias-clave-del-vendedor-40-jos%C3%A9gil>.

# The Impact of Tacit Knowledge Acquisition on Firm Performance: The Moderating Effect of Redundant Resources

Jing LIU,<sup>1</sup> and Fei YU

*Guilin University of Electronic Science and technology, China*

**Abstract.** With the advent of knowledge economy, any enterprise can be regarded as providing "knowledge service" for customers. Through the process of tacit knowledge acquisition and innovation, enterprises will eventually transform tacit knowledge into the core competitiveness of enterprises. Based on the sample data of Chinese listed manufacturing companies for three consecutive years, this paper analyzes the impact of three ways of tacit knowledge acquisition (namely: employee training, research and development and market research) on enterprise performance, and introduces the factor of redundant resources to analyze its moderating effect on this relationship. It is found that the three ways of tacit knowledge acquisition have significant positive effects on enterprise performance; different types of redundant resources (non precipitation redundant resources, precipitation redundant resources) strengthen and weaken the three positive relationship mechanisms. Based on the theory of resource-based view, it promotes the development process of knowledge acquisition and innovation under the knowledge economy of enterprises.

**Keyword.** Knowledge economy; tacit knowledge; acquisition method; redundant resources

## 1. Introduction

In 1999, with the rise of knowledge management theory research, many scholars at home and abroad began to invest in this field and published a series of articles. Effective knowledge management is an important way for enterprises to provide and continuously use knowledge to obtain competitive advantage [1]. Because tacit knowledge is implicit and difficult to imitate, it is not easy to be copied or stolen, and it is the source of enterprises to form core competitiveness. Compared with explicit knowledge, tacit knowledge can create more value, which is an important point in knowledge management [2].

How to obtain tacit knowledge from customers has become a hot topic in different industries and enterprises. The more secretive the knowledge is, the higher the value it brings, and enterprises will adopt a variety of technical methods to obtain tacit knowledge.

---

<sup>1</sup> Corresponding Author, Jing LIU, Guilin University of Electronic Science and technology, China;  
E-mail: 1106749360@qq.com

In addition, the acquisition of tacit knowledge is a resource consuming activity, especially for the tacit knowledge which is secret and difficult to manifest, it often needs the continuous resources; due to the uncertainty of the acquisition of tacit knowledge, some of the input resources may not bring any results to the enterprise. As an important "boost" of knowledge innovation, redundant resources are of great significance to tacit knowledge acquisition and innovation.

Many studies have found that tacit knowledge can bring high-value skills and experience, but now in this market with changeable demand, there are few quantitative studies on how to obtain tacit knowledge more effectively. Under the background of manufacturing industry, this paper analyzes three different ways of tacit knowledge acquisition, and explores the regulatory effect of redundant resources on each way of tacit knowledge acquisition.

## 2. Theoretical basis and research hypothesis

### 2.1 Enterprise tacit knowledge acquisition and enterprise performance

No matter the internal staff or the management, the external customers or the partners are the creators and storage of tacit knowledge. They are the origin of enterprise knowledge innovation. It is precisely because of the fluidity of knowledge that enterprises rely on the process of knowledge flow and transformation to realize knowledge innovation."Flow" includes one-way and multi-directional flow between tacit knowledge subjects. This process does not bring significant effect in tacit knowledge innovation, but it is a prerequisite for the process of tacit knowledge transformation.

Since the establishment of the enterprise, the knowledge flow and transformation of the enterprise is a continuous cycle process. The research and analysis of the "flow" of enterprise tacit knowledge has a certain impact on the successful "transformation" process, and the tacit knowledge innovation effect brought by the flow of different tacit knowledge carriers is different. It is necessary to grasp and analyze the complex relationship and law of the flow process.

Enterprises use the "flowing" tacit knowledge to improve their core competitiveness, so as to improve their performance [2]; secondly, enterprises have diversified tacit knowledge, which is conducive to enterprises to adapt to the dynamic environment, so as to make enterprises realize sustainable development; thirdly, after acquiring tacit knowledge, the main staff of enterprise goal implementation can improve their work efficiency and stimulate their knowledge innovation ability. Lu Xin et al. believe that tacit knowledge acquisition level has a positive impact on core competitiveness, clearly pointing out that the core competitiveness of enterprises has a positive relationship with the tacit knowledge acquisition level[3].Therefore, this paper believes that enterprise acquires tacit knowledge is beneficial to enterprise performance.

H1: tacit knowledge acquisition has a positive impact on firm performance.

Ren Zhijuan believes that the knowledge-based enterprise capability theory holds that the competitive advantage of enterprises does not come from knowledge itself, but from the acquisition and management of knowledge[4].According to Zhang Qingpu, tacit knowledge is distributed among employees, teams, departments, enterprises and other knowledge subjects at different levels [5].In the process of tacit knowledge flow,

tacit knowledge is generated by the self-renewal and mutual collision of these knowledge subjects. Starting from different knowledge subjects, there are different ways to acquire tacit knowledge.

Firstly, the tacit knowledge produced by individual employees. Employees have intuition, inspiration, observation, beliefs and other characteristics, which will produce tacit knowledge beyond the technical equipment, and can better meet the diverse needs of consumers. Through staff training and staff recruitment, new tacit knowledge can be brought to enterprises.

Due to the deeper understanding of the enterprise's products, and customers, the employees will produce a lot of tacit knowledge from their years of experience. If the enterprise trains them and grants them the right to learn skills, the employees will improve themselves in all aspects and summarize more innovative tacit knowledge; the formation of tacit knowledge will be accelerated. In addition, Lu Xin et al. believe that tacit knowledge acquisition of high-tech enterprises has a positive effect on organizational learning[5].

Second, tacit knowledge generated by groups (teams, departments).In a group, each knowledge individual has an environment of close interaction and direct communication. Through comprehension, absorption and transformation, tacit knowledge that can understand each other but can not be explained. Through the establishment of formal project development groups, as well as informal groups such as interest groups and communities.

Nowadays, more and more enterprises have set up R & D teams for a certain research project. The team members are slightly better than ordinary employees in terms of educational and work background. They have the skills of summarizing, transforming, analyzing and innovating knowledge. High-level personnel gather together and carry out ideological exchange, which can produce more tacit knowledge. Besides the satisfaction of soft resources, the R & D of marketable new products and services also requires the supply of hard resources such as technology and equipment.

Third, tacit knowledge produced by enterprises. This level can be divided into two different levels: The one is the tacit knowledge generated by the enterprise itself. In this stage, it requires continuous long-term integration and practice. The second is the tacit knowledge produced by the enterprise and other enterprises. With the help of "external brain", the enterprise introduces a variety of technical know-how and skills to exchange knowledge, and finally transforms the tacit knowledge into ones which is be mastered .This level can establish a wide range of knowledge alliances with customers, suppliers, universities, research institutions and even competitors[6].

In the past, in the process of independent R & D, the researchers of the enterprise only considered the design and development according to the product performance, aesthetics and other aspects, and purchased by customers. Now, if an enterprise wants to develop a successful new product, it must first have an in-depth understanding of customers' needs, which allows customers to participate in the development process of new products. In the process of two-way communication, after obtaining a large number of customer information needs, the enterprise finally splits the information knowledge, restructures and integrates new products. Therefore, we can think that after enterprises deepen the connection with the market and obtain a large amount of tacit knowledge from customers, they will promote enterprises to produce tacit knowledge related to new product development [7].

H2: there are significant differences in the impact of different ways of tacit knowledge acquisition on enterprise performance.

## 2.2 Regulatory role of redundant resources

Bourgeois Redundant resources are defined as the actual or potential idle resources in the organization, which can help the enterprise to cope with the pressure caused by internal and external changes[8]. Sufficient resources enable enterprises to selectively take actions to adapt to environmental changes, and provide support for the plans and projects , so as to improve enterprise performance. In all kinds of innovation activities, enterprises need to consume resources, and the innovation activity itself has great uncertainty. Even if enterprises have more redundant resources, it is difficult to meet all the resources required by innovation activities, and it is affected by the utilization efficiency of redundant resources.[8][9]

The literature in the field of tacit knowledge acquisition and innovation is less concerned with the regulatory role of redundant resources. Different types of resources combined with different ways of tacit knowledge acquisition will eventually produce different degrees of effect. According to the liquidity and flexibility of assets, redundant resources can be divided into two types: non precipitation and precipitation redundant resources. Therefore, this paper considers the different moderating effects of these two different resources on the relationship.[10]

### (1) Regulatory role of non precipitation redundant resources

Non precipitation redundancy refers to the resources with high liquidity and flexibility that are not designated for a specific purpose, including cash, reserve funds, etc., which can be temporarily transferred. In addition, this resources are easy to obtain, transfer and redistribute within the enterprise, and the consume fast. Due to the its traits, enterprise managers have greater autonomy on the resources, and even lead to more ineffective consumption. As a result, the education funds invested by enterprises for staff training actually become unprofitable management costs, such as the travel expenses and high training facilities reserve expenses, etc. Besides, employees are more likely to become dependent on this resource, and ignore their own non resource aspects of tacit knowledge creation skills. Therefore, we make conclude the following hypothesis:

H3: non precipitated redundant resources negatively regulate the relationship through employee training and improving enterprise performance.

Non precipitation redundancy is often used to help enterprises engage in new strategies, such as new product introduction and entering new markets. Because the allocation of this resource is easy to be unconstrained, enterprises can acquire new knowledge in a short time by investing them in new market development [11].Therefore, we can make the following hypothesis:

H4: non precipitated redundant resources positively regulate the relationship through market development and improving enterprise performance.

When facing new opportunities in the field of technology, managers of innovative enterprises can invest resources in time, through their own knowledge reserve and absorption and transformation ability of R & D personnel. Therefore, it can be concluded that: in the process of R & D, the use of non precipitation redundant resources can promote the acquisition of tacit knowledge, so as to achieve the improvement of enterprise performance.

H5: non precipitated redundant resources negatively regulate the relationship between enterprises acquiring tacit knowledge through R & D and improving enterprise performance.

(2) Regulatory role of precipitable redundant resources

Compared with non precipitation redundant resources, precipitation redundant resources have poor mobility and flexibility, which can only be directly applied to special situations. The use of precipitation redundant resources with poor flexibility will limit the method of generating new tacit knowledge, and employees cannot quickly and effectively allocate resources to the path that is conducive to their own acquisition, absorption and transformation of tacit knowledge, thus reducing the efficiency of tacit knowledge generated in employee training.

H6: precipitated redundant resources negatively regulate the relationship between tacit knowledge acquired through employee training and enterprise performance.

Because of its poor liquidity and flexibility, enterprises with more redundant resources are difficult to cope with the rapid changes of the external market. Facing the tacit knowledge of customers in the new market, managers can not respond in time and it is difficult to innovate enterprise tacit knowledge.

H7: precipitation of redundant resources negatively regulates the relationship between tacit knowledge acquired through market development and enterprise performance.

The characteristics of low liquidity can sometimes provide stable and lasting help for enterprises. Compared with the non-precipitation redundant resources, the greater the liquidity is, the re-transfer and redistribution within the enterprise will lead to poor stability. In addition, the use of the stable resources can reduce the risk of enterprise R & D projects, ensure the smooth progress of work, and further realize the purpose of obtaining tacit knowledge.

H8: precipitating redundant resources negatively regulates the relationship between enterprises acquiring tacit knowledge through R&D and improving enterprise performance.

### **3. Research design**

#### *3.1 Data collection*

Due to the incomplete data of the database in 2019, this paper selects the Chinese listed manufacturing industry( the industry classification of China Securities Regulatory Commission (CSRC) 2012 edition is used as the sample object) from 2016 to 2018 as the research sample, a total of 2161 enterprises, eliminating the companies with incomplete financial data and \* ST companies(A company with two consecutive years of special treatment for its losses), and finally obtains 50 effective enterprise samples. The original data are all from Chinese CSMAR database.

#### *3.2 Variable measurement*

The variable indexes involved in this paper come from the public notes to financial statements of Chinese listed companies. We use panel data to analyze the enterprises in different industries. Moreover, the relevant indicators are as follows.

(1) Dependent variable: This paper uses operating income to measure enterprise performance.

(2) Independent variable: The independent variables of this study are as follows:

Employee training: it mainly refers to the capital strength invested by the enterprise in the ability training and technical training of internal employees.

Research and development(R & D): refers to the financial support required by the enterprise R & D team to develop new products and services.

Market Research: the independent variable can be measured by the cost of all kinds of enterprises' involvement in the market.

(3) Regulatory variables: The regulating variable of this paper is redundant resources. Referring to Herold's measurement method, non precipitation resources are measured by dividing current assets by current liabilities. The precipitated redundant resources are measured by the owner's equity divided by the total liabilities formula.[12]

(4) Control variables

In order to ensure the normal distribution of data, this paper will use these three variables as control variables: enterprise size; enterprise age; industry.

## 4. Result analysis

### 4.1 Descriptive statistics and correlation coefficient analysis

By using the Eviews software can show the descriptive statistical results and correlation coefficient statistical results among the variables. And the control variables are significantly correlated with independent variables and dependent variables, which indicates that it is necessary to control the control variables. By calculating the variance inflation factor, it was found that the VIF values among all variables were all lower than 2, far less than the critical value of 10. Therefore, the problem of multicollinearity in the data of this sample was not serious.

Table 1 shows the results of multiple linear regression. Model 1 is a benchmark model with only control variables. In model 2, model 4 and Model 6, three independent variables were respectively added on the basis of the benchmark model, and their regression coefficients were significantly positive. The results show that the input cost of employee training, the input cost of market research and the input cost of R & D have different significant positive effects on enterprise performance, so hypothesis 1 and hypothesis 2 are verified.

In model 3, the variables of interaction between staff training and SR1, SR2 are added on the basis of the benchmark model, and the regression coefficient of employee training cost and two kinds of redundant resource interaction items is significantly negative, It shows that the use of these two redundant resources will weaken the positive relationship between tacit knowledge and enterprise performance. Hypothesis 3 and 6 are verified. In Model 5, hypothesis 4 can be verified by using the same analysis method in the previous paragraph; while the hypothesis 7 can not pass the test. In Model 7, hypothesis 5 can be verified; while the hypothesis 8 can not pass the test.

**Table 1** multiple regression model of tacit knowledge acquisition mode on Enterprise Performance

Independent variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Staff training	0.070315****		0.061843				
Marketing research				0.313445****	0.360398****		
R&D						0.078946***	0.168493**
Staff trainingxSR1			-0.154813*				
Staff trainingxSR2			-0.179038****				
Marketing researchxSR1					0.130779*		
Marketing researchxSR2					-0.068587		
R&D xSR1							0.144344*
R&D xSR2							0.021524

## 5. Research deficiencies and prospects

There are three limitations in this study: (1) Due to the limitation of selecting three independent variables at the same time for three consecutive years, only 50 enterprises are selected to meet the research purpose. Therefore, the follow-up study can consider using other indicators to measure the research object.(2) This research is an independent analysis and research through three levels of tacit knowledge subject, but does not research on enterprises taking a variety of tacit knowledge acquisition methods at the same time. In general, most enterprises can acquire tacit knowledge and innovate only in one way. Therefore, the follow-up research needs to deeply explore the interaction of different ways.(3) In this paper, the moderating effect of precipitation redundant resources is only considered for the three-year relationship between dependent variables and independent variables. In fact, the moderating effect of precipitation redundant resources may take a long time to reflect, and the lag data can be used for regression analysis.

## Reference

- [1] Martins V W B, Rampass I S, Anholon R, et al. Knowledge management in the context of sustainability: literature review and opportunities for future research [J].Journal of Cleaner Production, 2019,229: 489-500.
- [2] Yuan Qinghong, J enterprise intellectual capital management [M]. Beijing: economic management press, 2001.87-98
- [3] Lu Xin, Song Zaixun, Huang Jieyu.Research on the relationship between tacit knowledge acquisition level and core competence of high-tech enterprises [J]. Journal of Peking University of China (Social Science Edition),2019,35(06):108-114.
- [4] Ren Zhijuan. Research on the relationship between tacit knowledge acquisition, knowledge integration ability and firm innovation performance [D].Zhejiang Normal University,2018.
- [5] Zhang Qingpu, Li Zhichao. Characteristics and management of enterprise tacit knowledge [J]. Economic theory and economic management, 2002 (11): 47-50
- [6] Lu Xin, Huang Jieyu, Zhao Yun.Research on the influence of innovation climate on innovation performance in high-tech enterprises: the dual mediating role of implicit knowledge acquisition and organizational learning [J].Technological Economics and Management Research,2020(08):49-53.
- [7] Brown S L, Eisenhardt K M. Product development:Past research, present findings, and future directions[J].Academy of Management Review, 1995,20(2):343-378.
- [8] Bourgeois, L.J. On the measurement of organizational slack [J]. Academy of management Review, 1981, 6 (1): 29- 39

- [9] Sun Aiying, Su Zhongfeng. Research on the influence of resource redundancy on the choice of enterprise technological innovation [J]. Science of science and technology management, 2008 (05): 60-63 + 68
- [10] Jiang Weiping, Liu Daidi. Research on the relationship between R & D investment, redundant resources and enterprise performance [J]. Financial theory and practice, 2016,37 (05): 57-62
- [11] Zhang Qinglei, Shi Jianjun, Liu Chunlin. Research on the relationship between technological diversification, redundant resources and enterprise performance [J]. Scientific research management, 2015,36 (11): 21-28
- [12] Herold, D. M., Jayaraman, N., Narayanaswamy, C. R.. What Is the Relationship between Organizational Slack and Innovation?[J].Journal of Managerial Issues, 2006, 18(3): 372-392.

# Opinion-Aware Retrieval Models Based on Sentiment and Intensity of Lexical Features

Mohammad BAHRANI <sup>a,1</sup>, Thomas ROELKE <sup>a</sup>

<sup>a</sup>*School of Electronic Engineering and Computer Science, Queen Mary University of London*

**Abstract.** Sentiment analysis has received much attention in Information Retrieval (IR) and other domains including data mining, machine learning algorithms and NLP. However, when it comes to big data, incorporating sentiment of words into IR models becomes even more important, and as yet no widely accepted standard exists for this task. The contribution of this paper is a framework for quantifying term frequency (TF) variants with sentiments. We propose models derived from the strength of lexical features to improve sentiment-based ranking.

**Keywords.** Search and Analytics, Information Retrieval, Sentiment Analysis, Modern Ranking Models

## 1. Introduction and related work

Companies need to analyse customer's general feelings about their products. On the other hand, singular buyers want to know the sentiment of the product reviews before buying [1]. Wherefore the examination of sentiments would be beneficial for many applications specially in the field of big data. Researchers need to analyse the sentiment intensity over time to know about changes in rhetoric [2]. This would benefit analysts in companies, government and political departments that need to track emotions and attitudes. To date, sentiment analysis is mostly applied to the polarity (positive or negative) classification task. However, this may not be sufficient for many domains. Companies need to provide buyers with search engines that are able to retrieve top products based on user queries and sentiment analysis of reviews.

This paper proposes the research in the development of intensity-aware retrieval models in a generalizable framework. Sentiment analysis explores texts containing people's opinions, emotions and attitudes [3]. IR models take features such as term rareness, e.g. Inverse Document Frequency (IDF), into consideration to rank documents. However, they do not capture opinions in the retrieval process. For example, concerning movie reviews, the word '*good*' might occur nearly in every review, and from an IDF-point of view, it is not informative and selective. We expect a query such as '*good comedy movie*' to find

---

<sup>1</sup>Corresponding Author: Mohammad Bahrani, the school of Electronic Engineering and Computer Science, Queen Mary University of London, UK, Email address: m.bahrani@qmul.ac.uk.

good movies, but IR might process it similar to '*comedy movie*' due to the high frequency of the term *good* in the collection.

The main contribution of this paper is to address this problem by developing and investigating a sentiment-based framework for ranking reviews. This framework incorporates the sentiment and intensity of lexical features into IR units including the term frequency (TF) and IDF. Within this framework two different tasks are employed. Firstly, we add the IDF of sentiment-bearing words as a notion of rareness to the sentiment classification process. Secondly, we generalise IR models by proposing intensity-aware methods which take sentiment intensity into consideration. The idea is to regulate the term frequency by boosting weights of sentiment-bearing words. Such boosting is expected to overcome the problem caused by the rareness of these words with respect to IR models. Foundations of IR models are expressed in [4]. Additionally, opinion-based retrieval has been studied in a range of publications including [5,6,7,8]. Work related to our approach considers the use of semantics e.g. lexicons in machine learning and IR. [9] discussed the importance of aggregating semantics in IR models. [10] investigated the distribution of the emotions within textbooks, which resulted in the development of a framework for sentiment classification (SenticNet). [11] proposed a sentiment-aware attention method which leverages a three-step strategy to boost the performance sentiment analysis regarding movie reviews. Furthermore, [12] used IR techniques to classify sentences by polarity. Recently sentiment-aware approached received attention in deep learning [13]. An example of these applications is the incorporation of sentiment lexicons into neural networks [14]. VADER (Valence Aware Dictionary for sEntiment Reasoning) is a parsimonious rule-based tool for sentiment analysis concerning social media texts developed by Hutto and Gilbert [15]. It leverages a combination of qualitative and empirical methods using human experts and judgmental evaluations. Moreover, it employs a rich intensity-based lexicon to assign sentiments to sentences.

## 2. Opinion-Aware TF

### 2.1. Opinion-aware $TF_{total}$ :

We introduce micro and macro models for sentiment-based retrieval. These models are built upon VADER scores. Based on the VADER lexicon, a lexical feature could have a score between +1 and -1. Equation (1) shows how VADER calculates the compound sentiment of a sentence:

$$\text{sentiment}_{\text{vader}}(s) := \frac{\sum_{t \in s} \left( \sum_{i=1}^{n_L(t,s)} W_{\text{sentiment}}(t, i, s) \right)}{\sqrt{\left( \sum_{t \in s} \left( \sum_{i=1}^{n_L(t,s)} W_{\text{sentiment}}(t, i, s) \right) \right)^2} + \alpha} \quad (1)$$

$W_{\text{sentiment}}(t, i, s)$  is the sentiment score of the  $i_{th}$  occurrence of term  $t$ , and  $s$  is a sentence, and  $\alpha$  is a normalization parameter. The algorithm is reinforced by five heuristics including punctuation marks, capitalization, modifiers, negation and 'but' checker. It regulates the score of a word depending on the distance between the word and its degree modifier. The primary sentiment of a term is transformed into a sentence-dependent sentiment  $W_{\text{sentiment}}(t, i, s) := \bar{W}_{\text{sentiment}}(t) + \text{seed}(i, s)$ , where  $\text{seed}(i, s) := \sum_{j=0}^{i-1} \alpha(j, i, s)$ . For  $\alpha$ , the cases are:

$$\alpha(j, i, s) := \begin{cases} W_f(t_j(s), i - j) & \text{if } t_j(s) \text{ is an influencer (e.g. modifier)} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$\hat{W}_{\text{sentiment}}(t)$  is the primary sentence-independent sentiment of term  $t$ , and  $\text{seed}(t, s)$  is a weight to be added to this score in order to bring the heuristics into consideration.  $W_f$  estimates the weight of a single heuristic parameter in relation to the term by considering the distance and the constant weight of the parameter which is defined by VADER.

The intensity of a lexical feature is the sum of the absolute value of the sentiment and its corresponding seed weight  $W_{\text{intense}}(t, i, s) := |\hat{W}_{\text{sentiment}}(t)| + \text{seed}(i, s)$ . Subsequently, to rank the documents using VADER, we use  $RSV_{\text{vader}}(d) := \sum_{s \in d} \text{sentiment}_{\text{vader}}(s)$ .

The RSV shall be high if the document contains many positive opinion words. This is the rationale for the following opinion-aware TF variants:

### Definition 1 (Opinion-Aware Total TF Variants)

Let  $d$  be a document,  $s$  a sentence, and  $t$  a term (word). Let  $n_L(t, s)$  be the number of locations (positions) at which term  $t$  occurs in sentence  $s$ ; the notation  $n_L(t, d)$  applies to a document rather than a sentence.

$$TF_{\text{total-sentiment-Macro}}(t, d) := \sum_{s \in d} \left( \sum_{i=1}^{n_L(t, s)} W_{\text{sentiment}}(t, i, s) \right) \quad (3)$$

$$TF_{\text{total-sentiment-Micro}}(t, d) := n_L(t, d) \cdot \hat{W}_{\text{sentiment}}(t) \quad (4)$$

$$TF_{\text{total-intense}}(t, d) := \sum_{s \in d} \left( \sum_{i=1}^{n_L(t, s)} W_{\text{intense}}(t, i, s) \right) \quad (5)$$

Equation (3) shows the sentiment-based macro-TF  $TF_{\text{total}}(t, d)$ , where  $n_L(t, s)$  is the number of locations in which term  $t$  appears in sentence  $s$  and  $W_{\text{sentiment}}(t, i, s)$  returns the VADER score of the  $i_{th}$  occurrence of term  $t$  in the sentence.

In micro  $TF_{\text{total}}$ , Equation (4), we determine the term frequency independently and then multiply the result by the corresponding primary sentiment. Therefore, this model does not consider the impact of degree modifiers.

The opinion-aware TF can also be adopted from intensity or force of lexical features. In this paper, we considered the combination of strength and the corresponding seeds to determine the intensity weight  $W_{\text{intense}}(t, i, s)$  as expressed in Equation (5).

## 2.2. Opinion-Aware $TF_{BM25}$ :

A pivoted term frequency has been shown consistently to be important for retrieval (for the BM25 retrieval model, which can be viewed as a particular TF-IDF model). Therefore, we need pivoted term frequencies that are built upon sentiments in order to obtain a notion of the opinion-aware  $TF_{BM25}$ . The following definition lists new pivoted tf variants (lower-case tf indicates the pivoted TF variants).

## Definition 2 (Pivoted term frequencies)

$$\text{tf}_{\text{piv,sentiment-Macro}}(t, d) := \frac{\text{TF}_{\text{total-sentiment-Macro}}(t, d)}{(k_1(b \cdot \text{pivdl} + 1 - b))} \quad (6)$$

$$\text{tf}_{\text{piv,sentiment-Micro}}(t, d) := \text{tf}_{\text{piv}}(t, d) \cdot \hat{W}_{\text{sentiment}}(t) \quad (7)$$

$$\text{tf}_{\text{piv,intense}}(t, d) := \frac{\text{TF}_{\text{total-intense}}(t, d)}{(k_1(b \cdot \text{pivdl} + 1 - b))} \quad (8)$$

The rationale is as follows. A scaling of the total TF  $\text{tf}(t, d)$  is not advisable since this would have implications on the document length. A scaling of the  $\text{TF}_{\text{BM25}}$  would just equate to a linear scaling of the TF.IDF weight. The determination of the opinion-aware  $\text{TF}_{\text{BM25}}$  is presented in the following definition.

## Definition 3 (Opinion-Aware $\text{TF}_{\text{BM25}}$ Variants)

$$\text{TF}_{\text{BM25-sentiment-Macro}}(t, d) := \frac{\text{tf}_{\text{piv,sentiment-Macro}}(t, d)}{|\text{tf}_{\text{piv,sentiment-Macro}}(t, d)| + 1} \quad (9)$$

$$\text{TF}_{\text{BM25-sentiment-Micro}}(t, d) := \frac{\text{tf}_{\text{piv,sentiment-Micro}}(t, d)}{|\text{tf}_{\text{piv,sentiment-Micro}}(t, d)| + 1} \quad (10)$$

$$\text{TF}_{\text{BM25-intense}}(t, d) := \frac{\text{tf}_{\text{piv,intense}}(t, d)}{|\text{tf}_{\text{piv,intense}}(t, d)| + 1} \quad (11)$$

For the scientific study of intensity, in this paper we focused on opinion words. All of the proposed models consider the neutral terms within queries as stop-word (words to be ignored). However, future studies are needed to explore the integration of topical retrieval with opinion words.

## 3. Proposed Models

### 3.1. Term-Frequency – Inverse-Document-Frequency (TF.IDF)

The proposed TF.IDF consists of the well-known IDF as the notion of rareness and opinion-based term frequencies  $W_{\text{TF.IDF}-x}(t, d) := \text{TF}_x(t, d) \cdot \text{IDF}(t)$ .  $x$  is a generic type which can be any of different forms of opinion-aware approaches including total – sentiment – Macro, total – sentiment – Micro, total – intense and the corresponding BM25 approaches. The Retrieval Status Value (RSV) is the sum of TF.IDF weights across document terms  $\text{RSV}_{\text{TF.IDF}-x}(d, c) := \sum_{t \in d} W_{\text{TF.IDF}-x}(t, d)$ .

### 3.2. Language Modelling (LM)

For LM, we need an approach that considers sentiment when estimating the within-document term probability  $p(t|d)$  and the collection-wide term probability  $p(t|c)$ , respectively. For reflecting the fact that we apply negative values (because of the polarity), we introduce the notation  $\pi(t|d)$  and  $\pi(t|c)$ . We hire opinion-aware term frequencies introduced in the previous section and incorporate them into the probabilities. Therefore, we determine the new parameters as follows:

$$\pi_x(t|d) := \frac{\text{TF}_{\text{total}-x}(t,d)}{|d_x|} \quad (12)$$

$$\pi_x(t|c) := \frac{\text{TF}_{\text{total}-x}(t,c)}{|c_x|} \quad (13)$$

$x$  is the type of opinion-aware model,  $d_x$  is opinion-based document length and  $c_x$  denotes collection length. The determination of the length parameters is dependent on the used model-type  $x$ . The calculation of opinion-aware LM would result in issues related to negative values within logarithm. To address this issue, we apply the logarithm to the absolute result of the division and deliver the polarity of term sentiment into the formula by the use of  $\text{TF}_x(t,q)$  parameter which determines the sign. In LM, the TF quantification is for the *query*; in other words, on the TF-IDF side of IR, it is more straight-forward to generalise the TF regarding sentiment. Below we show the opinion-aware LM:

$$W_{LM-x}(t,d,q,c) := \text{TF}_x(t,q) \cdot \log \left( \left| \frac{(1 - \sigma_d) \cdot \pi_x(t|c) + \sigma_d \cdot \pi_x(t|d))}{\pi_x(t|c)} \right| \right) \quad (14)$$

The document is ranked by dividing the smoothed version of the multinomial probability of the query given the document by the probability of the query in the collection. Therefore, the corresponding RSV is defined as  $RSV_{LM-x}(d,q,c) := \sum_{t \in q} W_{LM-x}(t,d,q,c)$ .

## 4. Experiments

The basic models process all of query-terms regardless of the sentiments, whereas the intensity-aware models only consider the intensity values of sentiment-bearing terms within queries and ignore neutral words. To perform the evaluation task, we used the IMDB-review collection as the primary dataset. For confirming the results, we used all 2000 DVD reviews exists in the Amazon Sentiment Dataset [16]. Each query set consists of 50 reviews which are correspondent to the query set label in terms of polarity. As an example, '*There are scenes which make you gulp with sudden emotion, and those which even put a smile on your face through ...*' is a snippet of a positive query that we used. To evaluate the intensity-aware models and their corresponding basic models, we hired Mean Average Precision (MAP) and Reciprocal Rank as shown in table 1. All of the novel models provided higher MAP scores than the basic models for both query sets. The models were more effective for negative reviews than the positive ones concerning IMDB dataset whereas, they provided much higher scores for the positive set compared to the negative queries when applied on Amazon DVD reviews.  $LM_{intense}$  and  $TF.IDF_{BM25-intense}$  achieved the highest MAP and Reciprocal Rank values. Although the intensity-aware LM provided a higher MAP score than the macro version of the  $TF.IDF_{BM25}$ , the variance is extremely small.

we performed a query-based analysis to capture the statistics of queries which are more compatible with the novel models for both positive and negative query sets.

We calculated the average of APs for both basic and intensity-aware models separately and subsequently for each query set, plotted the distribution of the differences ordered by descending. As can be seen in Figure 1, the intensity aware models were more effective

**Table 1.** Evaluation of Intensity-Aware Retrieval Models: Intense methods worked better than their corresponding basic models.

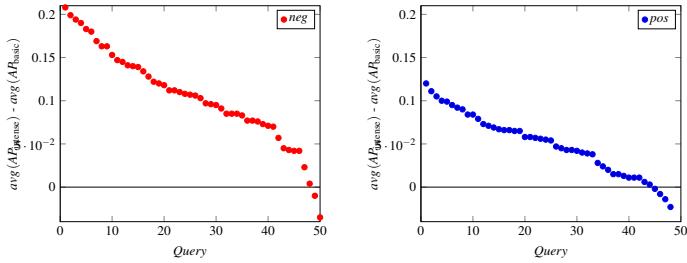
Model	MAP			Reciprocal Rank
	pos	neg	avg	total
<b>IMDB Reviews</b>				
TF.IDF <sub>total</sub> -basic	0.286	0.237	0.261	0.80
TF.IDF <sub>total</sub> -intense	0.320	0.345	0.332	(+27.2%)
TF.IDF <sub>BM25</sub> -basic	0.293	0.234	0.263	0.87
TF.IDF <sub>BM25</sub> -intense	0.351	0.322	<u>0.336</u>	(+27.76%)
LM <sub>basic</sub>	0.259	0.241	0.250	0.60
LM <sub>intense</sub>	0.305	0.369	<u>0.337</u>	(+34.8%)
<b>Amazon DVD Reviews</b>				
TF.IDF <sub>total</sub> -basic	0.288	0.111	0.199	0.826
TF.IDF <sub>total</sub> -intense	0.478	0.115	0.296	(+48.74%)
TF.IDF <sub>BM25</sub> -basic	0.280	0.125	0.202	0.886
TF.IDF <sub>BM25</sub> -intense	0.471	0.138	<u>0.304</u>	(+50.49%)
LM <sub>basic</sub>	0.281	0.090	0.185	0.750
LM <sub>intense</sub>	0.465	0.145	<u>0.305</u>	(+64.86%)
				<u>0.995</u>

**Table 2.** Sentiment Polarity Classification: Sentiment models had higher F1 scores than the baseline. Sentiment-aware macro TF.IDF<sub>BM25</sub> was more effective than other models.

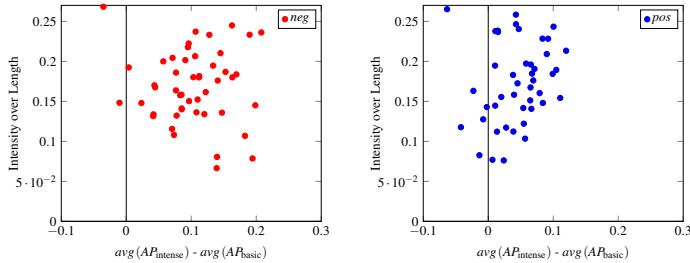
Model	P@1000		R@1000		F1	
	pos	neg	pos	neg	pos	neg
<b>IMDB Reviews</b>						
sentiment <sub>vader</sub>	0.893	0.724	0.0714	0.0579	0.132	0.107
TF.IDF <sub>total</sub> -sentiment-Macro	0.905	0.764	0.0724	0.0611	0.134 (+1.5%)	0.113 (+5.6%)
TF.IDF <sub>total</sub> -sentiment-Micro	0.898	0.762	0.0718	0.0610	0.132 (+0.0%)	0.112 (+4.7%)
TF.IDF <sub>BM25</sub> -sentiment-Macro	0.953	0.804	0.0762	0.0643	<u>0.141</u> (+6.8%)	<u>0.119</u> (+11.2%)
TF.IDF <sub>BM25</sub> -sentiment-Micro	0.943	0.801	0.0754	0.0641	0.139 (+5.3%)	0.118 (+10.3%)
LM <sub>sentiment</sub> -Macro	0.926	0.773	0.0741	0.0618	0.137 (+3.8%)	0.114 (+6.5%)
LM <sub>sentiment</sub> -Micro	0.919	0.780	0.0735	0.0624	0.136 (+3.0%)	0.115 (+7.5%)

on more than 95% of the queries compared to the basic approach. Figure 2 shows the positive correlation between the quality of the intensity-aware models and the ratio of the query intensity to the query length. Interestingly, the correlation is stronger for positive queries which shows that the polarity of the queries could impact effectiveness of the models. The Pearson correlation coefficient for positive query set is 0.21 and for the negative query set is estimated as weak positive.

Moreover, we applied the plain sentiment-based VADER as well as the proposed sentiment-aware IR models on the IMDB dataset containing 25000 highly polar reviews [17]. The dataset is divided equally into negative and positive parts. We used the data and their labels as the gold standard for this task. Top 1000 reviews retrieved by models are labelled as *pos* and accordingly top 1000 reviews from the bottom of the reversed



**Figure 1.** Distribution of Avg-AP Differences between Intense and Basic Models considering 100 Queries (in descending order): Query Analysis shows intense models were more effective for roughly 96% of queries compared to the basic models.



**Figure 2.** Pearson Correlation between Avg-AP Differences and the Ratio of Query-Intensity to Query Length: The correlation value for positive queries is 0.21 while the relationship between the parameters is weak positive regarding negative queries.

result-lists fell into the *neg* class. Table 2 column 4 lists F1 scores for our runs on the dataset. The data in this column indicates that all of the sentiment-aware models outperformed the baseline *sentiment<sub>vader</sub>* concerning both negative and positive classifications. The macro instance of the TF.IDF<sub>BM25</sub> achieved the highest score among the models. The advantage of the macro group is the consideration of influencers in retrieval. As we expected, they provided higher scores than the micro models although the differences are not statistically significant. In contrary with the baseline, the performance of the novel models was higher for the negative class compared to the positive one.

## 5. Conclusion and Future Work

In this paper, we presented two novel families of opinion-aware models, namely sentiment-aware and intensity-aware models to deal with the problem of the opinion words with low IDF and high intensity. This study also explored the consideration of a notion of IDF in sentiment classifications. To investigate the use of sentiment intensity in retrieval, we applied both basic and intensity-aware models to movie reviews and tested to find out if items of a specific polarity retrieve similar items. All of the intensity-aware models outperformed their corresponding basic models. It turned out that the effectiveness of the novel models is consistent across a range of positive and negative queries. The deficiency of the approach is a lack of advanced natural language understanding. Future work could improve the proposed framework by applying a Natural Language Pro-

cessing (NLP) tool that takes into consideration synonyms, antonyms and slang phrases. Moreover, further studies are required to enhance the quality of the determination LM parameters. This paper paves the way to achieving standards for considering sentiment in data. Those standards will be important for the many applications that process big data.

## References

- [1] Wawre SV, Deshmukh SN. Sentiment classification using machine learning techniques. *International Journal of Science and Research (IJSR)*. 2016;5(4):819–821.
- [2] Wilson T, Wiebe J, Hwa R. Just how mad are you? Finding strong and weak opinion clauses. In: aaai. vol. 4; 2004. p. 761–769.
- [3] Liu B. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*. 2012;5(1):1–167.
- [4] Roelleke T. Information retrieval models: foundations and relationships. *Synthesis Lectures on Information Concepts, Retrieval, and Services*. 2013;5(3):1–163.
- [5] Zhang W, Yu CT, Meng W. Opinion retrieval from blogs. In: Silva MJ, Laender AHF, Baeza-Yates RA, McGuinness DL, Olstad B, Olsen ØH, et al., editors. *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management, CIKM 2007, Lisbon, Portugal, November 6-10, 2007*. ACM; 2007. p. 831–840. Available from: <https://doi.org/10.1145/1321440.1321555>.
- [6] Gerani S, Carman MJ, Crestani F. Proximity-based opinion retrieval. In: *Proceedings of the conference on research and development in information retrieval*. ACM; 2010. .
- [7] He B, Macdonald C, He J, Ounis I. An effective statistical approach to blog post opinion retrieval. In: *Proceedings of the 17th ACM conference on Information and knowledge management*. ACM; 2008. p. 1063–1072.
- [8] Huang X, Croft WB. A unified relevance model for opinion retrieval. In: *Proceedings of the 18th ACM conference on Information and knowledge management*. ACM; 2009. p. 947–956.
- [9] Bahrani M, Roelleke T. FDCM: Towards Balanced and Generalizable Concept-based Models for Effective Medical Ranking. In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*; 2020. p. 1957–1960.
- [10] Bisio F, Meda C, Gastaldo P, Zunino R, Cambria E. Sentiment-oriented information retrieval: Affective analysis of documents based on the senticnet framework. In: *Sentiment analysis and ontology engineering*. Springer; 2016. p. 175–197.
- [11] Lei Z, Yang Y, Yang M. SAAN: A sentiment-aware attention network for sentiment analysis. In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*; 2018. p. 1197–1200.
- [12] Dragoni M. Shellfbk: an information retrieval-based system for multi-domain sentiment analysis. In: *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*; 2015. p. 502–509.
- [13] Zhang L, Wang S, Liu B. Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 2018;8(4):e1253.
- [14] Wu C, Wu F, Liu J, Huang Y, Xie X. Sentiment Lexicon Enhanced Neural Sentiment Classification. In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*; 2019. p. 1091–1100.
- [15] Hutto CJ, Gilbert E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: *Eighth international AAAI conference on weblogs and social media*; 2014. .
- [16] Blitzer J, Dredze M, Pereira F. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In: *Proceedings of the 45th annual meeting of the association of computational linguistics*; 2007. p. 440–447.
- [17] Maas AL, Daly RE, Pham PT, Huang D, Ng AY, Potts C. Learning Word Vectors for Sentiment Analysis. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Portland, Oregon, USA: Association for Computational Linguistics; 2011. p. 142–150. Available from: <http://www.aclweb.org/anthology/P11-1015>.

# Financial Flexibility During the Pre- and Post-Global Financial Crisis Periods

Weihan Cui<sup>a,1</sup>

<sup>a</sup>*Department of Economics, Nagoya University of Economics, Japan*

**Abstract.** This paper investigates the determinants of financial flexibility of Japanese firms before and after the global financial crisis. In the pre-crisis period, growth opportunity has a positive effect on financial flexibility, but in the post-crisis period, this effect turns to negative and is especially strong for financially constrained firms. These results indicate that in normal time, the Japanese firms pursue financial flexibility for investment demand as argued in previous literature. During difficult time, investment environment deteriorates and investment declines, therefore the firms with less growth opportunity accelerate to accumulate financial flexibility, especially for financially constrained firms as they suffer more than others in such a period. Enhancing the investment environment can improve the efficiency of corporate capital and financial support from banking system may ease the stress of financially constrained firms in post-crisis period.

**Keywords.** Financial flexibility; Debt conservatism; Global financial crisis

## 1. Introduction

Financial flexibility refers to a certain ability of a firm, which enables it reacts to investment demand or crisis in a timely and necessarily way. After the 2008 Global Financial Crisis, acquiring and maintaining financial flexibility have become increasingly important for firms. Firms pursue financial flexibility by holding enormous cash holdings or debt capacity, which means accumulating cash or decreasing leverage. A substantial number of firms choose to hold low or even zero-leverage [1] [2], which also known as conservative debt policy [3]. These conservative debt policies can provide firms with the capability of issuing debt when the necessity arises, and therefore provide financial flexibility.

Due to the strong bank-firm relationship, though [4] argue that zero-leverage is an international phenomenon, Japan has extremely low zero-leverage firm ratio. [5] analyzes Japanese zero-leverage firms and concludes that Japanese firms are kept away from zero-leverage by enormous bank power. Having positive gross leverage does not mean that Japanese firms are not conservative, in fact, Japanese firms choose to simultaneously have positive gross debt and sufficient cash holdings which is enough to repay all the debt. This behavior results in a positive gross debt ratio and a negative net debt ratio, and this so-called “non-positive debt policy” becomes the conservative

---

<sup>1</sup> Corresponding Author, Weihan Cui, Department of Economics, Nagoya University of Economics, 61-1 Uchikubo, Inuyama City, Aichi Prefecture 484-8504, Japan; Email address: cui-w@nagoya-ku.ac.jp. This research is funded by Yamada Memorial Trust Fund for Promotion of Scholarly Research.

debt policy of Japanese firms. By the end of 2017, 59% of Japanese firms are non-positive net debt firms (the data comes from Nikkei), the non-positive debt policy is truly the main stream in Japan.

Though the determinants of conservative debt policy are still unclear, the previous literature does have one consensus: the firms with more growth opportunity are more likely to adopt conservative debt policy [1] [4] [6] [7] [8] [9] [10]. The empirical results show that firms with higher market-to-book ratio (which is the proxy of growth opportunity) are more likely to adopt conservative debt policy, which implies that firms save debt capacity and accumulate financial flexibility for valuable investment opportunity in the future.

Because conservative debt policies provide financial flexibility, the Japanese firms with non-positive net debt can also be regarded as firms with considerable financial flexibility. If Japanese firms accumulate financial flexibility for future investment opportunity as previous literature argued, a positive effect of growth opportunity on the possibility of non-positive net debt is expected. However, this effect is negative for Japanese firms [11]. This evidence indicates that the lack of valuable investment opportunity limits the use of funds, meanwhile, Japanese firms resume to accumulate cash holdings in recent years, which together results in an increase of non-positive net debt firms.

During and after the global financial crisis in 2008, external credit supply becomes constrained and financial flexibility becomes more valuable than normal times. Since the number of Japanese non-positive net debt firms dramatically increased right after the global financial crisis, it is natural to consider that the financial crisis is one of the determinants of financial flexibility. This paper analyzes the determinants of financial flexibility of Japanese firms before and after the global financial crisis. The results suggest that, in pre-crisis period, there exists a positive effect of growth opportunity on the financial flexibility, which is in line with the most previous literature. However, this positive effect turns to negative in the post-crisis period and gradually weakened as time goes by. Furthermore, the negative effect of growth opportunity is extremely strong for financially constrained firms before the crisis and gradually weakened afterward. These results indicate that there are different motivations for accumulating financial flexibility based on the macro environment. In normal time, firms accumulate cash holdings to prepare for the future investment opportunity; in difficult time, investment environment deteriorates and firms with less growth opportunity accumulate financial flexibility naturally due to decreased investment. Meanwhile, financially constrained firms suffer more than others and are more urgent to save financial flexibility in the post-crisis period.

This paper present empirical evidence which contributes to the understanding of determinants of financial flexibility, while previous literature focus on the positive relation between growth opportunity and financial flexibility, this paper provides results which show that there exits different growth opportunity-financial flexibility relation under different macro environment. This paper also contributes to the understanding of Japanese conservative debt policy, especially the determinants of non-positive net debt of Japanese firms around the global financial crisis, providing possible explanation for the existing empirical results of Japanese conservative debt policy that differ from the other countries.

The remainder of this paper is organized as follows. Section 2 explains the main hypothesis and econometric methodology. Section 3 presents the main results, and Section 4 concludes the paper.

## 2. Hypotheses and Methodology

With market friction's existence, firms eschew debt to mitigate investment distortions and prepare for valuable future investments or unexpected cash shortfalls [12]. [8] study the firms that persistently maintain low leverages and find that such firms are highly profitable, generate sufficient cash flows, have higher growth opportunity, and are less likely to be financially constrained. As substantial previous literature agree that firms with more growth opportunity are more likely to have lower leverage, in other words, more debt capacity or financial flexibility.

While U.S firms take the equity market as their prior capital resource, Japanese firms mainly rely on banks for borrowing [13]. The relationship with their "main bank" is important, because this bank relationship is not only regarded as access to external funds, but also (to some extent) plays a role as a guarantee of a firm's performance; those firms which do not have a main bank may be considered as not bank-relationship-worthy. Furthermore, the bank-firm relationship is expensive to recover if terminated [5], so the Japanese firms are reluctant to repay all the debt. Therefore, holding more cash than debt becomes the conservative debt policy in Japan. As Figure 1 shows, the number of Non-positive net debt (NPND) firms has been generally increased in recent years.



Figure 1. Trend of Non-positive Net Debt Firms in Japan.

The previous literature adopts various proxies of financial flexibility, such as the market value of the real estate, short-term debt, cash holdings, and excess cash. Given the prevalence of NPND firms in Japan, this paper adopts two proxies of financial flexibility. One proxy is a dummy variable of NPND, which takes the value of one if a firm has non-positive net debt and zero otherwise; the other proxy is excess cash holdings, defined as the cash holding minus the debt.

A positive shock to firms' financial flexibility will decrease the value of cash and therefore the amount of cash holdings, especially for financially constrained firms. The global financial crisis is associated with a negative demand shock [14] and caused a decline in growth opportunity [15]. Therefore, the value of cash and the amount of cash holdings are supposed to increase after the crisis. Meanwhile, Japan is known for low level of growth opportunity, lack of growth opportunity encourages the stockpiling of cash holding. Since that exists a positive relationship between growth opportunity and leverage for low growth firms [16], with other things being equal, a lower leverage means a higher possibility of having NPND. Since the previous literature suggests that the effect of growth opportunity on NPND is negative [11], it is possible that this relationship is stronger in post-crisis period than pre-crisis period.

Hypothesis 1: There exists a negative relationship between growth opportunity and the possibility of accumulating financial flexibility, is stronger in post-crisis period than pre-crisis period.

Furthermore, financially constrained firms value cash more than others [17] [18]. Firms with poor cash holdings before the crisis are likely to experience more severe investment decline [19], and financially constrained firms are also more likely to have payout reduction in the post-crisis period [15]. During the financial crisis, financially constrained firms are more sensitive to the deteriorated environment and are more urgent to accumulate cash holdings. Therefore, the negative relationship between growth opportunity and NPND could be stronger for financially constrained firms.

Hypothesis 2: The negative relation between growth opportunity and the possibility of accumulating financial flexibility is especially strong for financially constrained firms.

The original data consist of the firms which are contained in Nikkei NEEDS-Financial QUEST data set during the period from 1983 to 2015. The global financial crisis period is set from 2005 to 2012. The pre-crisis period is defined as from 2005 to 2007; the post-crisis period is defined as from 2008 to 2010 and extended into 2008 to 2012 in the additional specifications.

**Table 1.** Statistic Summary

	2005~2010			2005~2012		
	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
Excess Cash	6085	-0.06	0.23	5730	-0.05	0.24
NPND	6085	0.32	0.46	5730	0.33	0.47
[M/B]	5884	1.23	0.70	5632	0.93	0.36
Size	5884	10.87	1.43	5632	10.87	1.49
Tangibility	5884	0.30	0.17	5632	0.31	0.18
Profitability	3261	0.00	0.00	2820	0.00	0.00
Dividend	5275	0.01	0.01	5444	0.01	0.02
[R&D]	4221	0.03	0.15	4081	0.03	0.13
Age	5884	2.31	0.74	5632	2.53	0.67
CAPEX	5884	0.01	0.08	5632	-0.02	0.14

Excess cash is defined as cash holding + securities in liquid asset – short-term debt – long-term debt – bond, scaled by total asset. NPND is a dummy variable that represents the non-positive net debt position, which equals one if a firm has non-positive net debt and zero otherwise. [M/B] is the ratio of the market asset to the book asset. Size is the natural log of the total asset. Tangibility is defined as the tangible asset divided by the total asset. Profitability is defined as the sum of income and depreciation, divided by the total asset. Dividend is the ratio of dividend to the total asset. [R&D] is defined as the ratio of [R&D] to sales. Age is the natural log of the number of years that a firm appears in the sample period. CAPEX represents the capital expenditure, defined as the difference of fixed asset divided by the total asset. The summary statistics are shown in Table 1.

### 3. Main Results

The analysis is conducted in two steps. First, following [20], the dependent variable NPND dummy which represents both conservative debt position and sufficient financial flexibility, is regressed on the main variables ([M/B], Size, Tangibility,

Profitability, Dividend) and control variables ([R&D], Age, CAPEX). Then, the dependent variable will be replaced by excess cash. The regressions are applied to fixed effect to control for the time-invariant variables. All independent variables are taken one lag to control the endogeneity issue.

Table 2 shows the main results of logistic regression for all firms, the dependent variable is the dummy variable NPND. Though insignificantly, the [M/B] is negative in 2005-2010, which is in the line with [11]. However, the sign of [M/B] is significantly positive in 2005-2007, which means that before the crisis, the firms with more growth opportunity are more likely to accumulate financial flexibility. This result is contrast to the [11], but consistent with the previous literature on low and zero-leverage.

**Table 2.** Panel analysis of NPND in Pre-and Post-financial crisis period

	(1)	(2)	(3)	(4)
VARIABLES	2005~2010	2005~2007	2008~2010	2008~2012
[M/B]	-0.250 (0.366)	1.827* (1.044)	-4.492** (1.906)	-1.571* (0.862)
Size	-3.558*** (1.016)	-3.182 (2.716)	-13.495*** (3.945)	-7.474*** (1.590)
Tangibility	-8.545** (3.744)	-1.496 (8.607)	-40.800*** (13.128)	-24.749*** (5.300)
Profitability	750.096* (428.166)	673.909 (915.963)	3,460.176 (2,296.441)	1,187.199 (826.360)
Dividend	86.451** (36.475)	147.449 (130.032)	-32.344 (61.047)	28.061 (35.426)
[R&D]	-38.069** (17.643)	-58.640 (53.733)	-43.228 (38.029)	-12.994 (19.143)
Age	2.182*** (0.583)	2.547 (1.948)	2.063 (1.841)	2.858*** (0.949)
CAPEX	2.437 (2.296)	4.889 (5.174)	8.459 (5.820)	1.467 (2.145)
Observations	560	156	174	509
Number of firms	125	56	64	125

In 2008-2010 period, the signs of [M/B] becomes strongly significant and negative. As the post-crisis timespan extended to 2008-2012, the negative effect of [M/B] weakens and the significance compromises a little. The hypothesis H1 is supported.

These results imply that the 2008 global financial crisis triggered a fierce negative effect of growth opportunity on the financial flexibility, but this motivation dries out as the time goes by. Though previous literature of Japanese conservative debt policy argue that Japanese firms accumulate financial flexibility because they have nowhere to invest, but these results suggest that, in normal time, Japanese firms' behavior is not different from the other countries' firms. They pursue financial flexibility and prepare for future investment opportunity, which is also referred as "precautionary motivation". One possible explanation for the negative relation between growth opportunity and conservative debt policy in [11] is that, the negative relation appears in the post-crisis period is so strong that it biased the result for the whole sample period. As time goes by, the impact of global financial crisis fades away and the positive relation revives, therefore the negative relation weakens gradually.

**Table 3.** Panel analysis of NPND in Post-financial crisis period for firms that have not been NPND in 2005~2007 (Column 1 and 2) and firms that have never been NPND before 2008 (Column 3 and 4)

	(1)	(2)	(3)	(4)
VARIABLES	2008~2010	2008~2012	2008~2010	2008~2012
[M/B]	-0.722* 0.110		-4.025** -0.412	

	(0.432)	(0.262)	(1.568)	(0.469)
Size	-3.552*** (0.754)	-4.438*** (0.769)	-4.459*** (1.301)	-5.645*** (1.356)
Tangibility	-17.974*** (4.044)	-16.811*** (3.266)	-14.517** (6.916)	-13.559** (5.269)
Profitability	325.385 (668.098)	-2.829 (124.541)	818.311 (922.226)	-20.628 (147.417)
Dividend	90.836** (41.399)	122.580*** (37.766)	170.403** (66.333)	159.448*** (57.548)
[R&D]	-7.665 (23.384)	-9.181 (14.471)	-76.259 (51.499)	-39.813 (41.430)
Age	2.825*** (0.444)	3.848*** (0.442)	5.106*** (1.000)	7.382*** (1.020)
CAPEX	-2.918 (2.344)	-2.715 (2.119)	-1.610 (4.582)	-3.556 (3.654)
Observations	546	877	332	547
Number of firms	57	90	35	57

To investigate the determinants of financial flexibility of firms with potential financial constraints, a specification of analysis that focuses on the newly-becoming NPND firms in the post-crisis period is conducted. The results are showed in Table 3. The same logistic regression in Table 2 is applied to the firms that (1) do not have NPND in the pre-crisis period i.e. 2005-2007 (Column 1 and 2), (2) do not have NPND in all years before the crisis i.e. before 2008 (Column 3 and 4). Again, there is a strong negative effect of [M/B] on financial flexibility in the post-crisis period, and this negative effect weakens as the timespan expanded. Meanwhile, the negative effect is strongest for firms that have not been NPND firms before the financial crisis, the hypothesis H2 is supported by these results.

If we take the absent record of NPND before the financial crisis as the sign of being financially constrained, these results can also be comprehended as follows: in the post-crisis period, growth opportunity has the strongest negative effect on financial flexibility for the most financially constrained firms. Such financially constrained firms are likely to accumulate cash holdings to mitigate financial frictions and save more aggressively when the situation deteriorates [21]. However, the strong effect of [M/B] may also arise from the desire for the market reputation, as the least evaluated firms in the equity market suffer more than others, such financially constrained firms may have stronger motivation to pursue financial flexibility to acquire market reputation. Meanwhile, during the global financial crisis, some of the Japanese firms experienced severe capital shortage. As Japan is a bank-centered economy, cannot borrow from bank means that firms need to rely on cash holdings. This may be another reason for the acceleration of accumulating financial flexibility of financially constrained firms after the crisis.

Though NPND firms are considered as financially flexible, there may exist huge differences in the excess cash between each firm. To further investigate the relation between the thickness of financial flexibility and firms' growth opportunity, another specification using the amount of excess cash as the dependent variable is conducted. The results are displayed in Table 4. Column (1) and (2) show the results of the whole timespan from 2005 to 2010, Column (3) and (4) show the pre-crisis period results (from 2005 to 2007), Column (5) and (6) show the results of the post-crisis period (from 2008 to 2010). For the timespan of 2005-2010, [M/B] ratio shows significantly positive signs, which implies a positive effect of growth opportunity on excess cash. This positive effect of growth opportunity on excess cash is stronger in the pre-crisis

period than in the whole timespan and disappears in the post-crisis period. These results suggest that, in the pre-crisis period, firms with more growth opportunity are more likely to accumulate financial flexibility. Though this effect disappears in the post-crisis period 2008-2010, the results of timespan 2005-2010 is significantly positive, which indicates that the positive effect before the crisis is sufficiently strong. These results reinforce the integrity of the corresponding results in Table 2.

**Table 4.** Panel analysis of financial flexibility in Pre-and Post-financial crisis period of all firms with fixed effect

VARIABLE	(1)	(2)	(3)	(4)	(5)	(6)
S [M/B]	2005~2010 0.011** (0.005)	2005~2010 0.017*** (0.005)	2005~2007 0.028*** (0.006)	2005~2007 0.027*** (0.006)	2008~2010 0.006 (0.013)	2008~2010 0.002 (0.013)
Size	-0.089*** (0.009)	-0.097*** (0.009)	-0.070*** (0.014)	-0.081*** (0.017)	-0.154*** (0.017)	-0.128*** (0.018)
Tangibility	-0.390*** (0.033)	-0.393*** (0.033)	-0.235*** (0.054)	-0.240*** (0.055)	-0.269*** (0.058)	-0.271*** (0.057)
Profitability	2.339*** (0.828)	2.366*** (0.824)	1.378 (1.001)	1.329 (1.001)	1.077 (1.687)	0.961 (1.631)
Dividend	0.175 (0.161)	0.120 (0.161)	2.672*** (0.798)	2.548*** (0.806)	-2.050*** (0.538)	-1.033* (0.544)
[R&D]		-0.584*** (0.197)		-0.667* (0.391)		-0.241 (0.288)
Age		0.036*** (0.007)		0.007 (0.014)		0.136*** (0.016)
CAPEX		0.038* (0.020)		0.022 (0.031)		0.072** (0.028)
Constant	0.955*** (0.102)	0.954*** (0.109)	0.658*** (0.160)	0.781*** (0.182)	1.643*** (0.194)	0.998*** (0.225)
Observations	3,716	3,716	1,845	1,845	1,871	1,871
R-squared	0.081	0.093	0.079	0.082	0.096	0.161
Number of firms	967	967	806	806	765	765

#### 4. Conclusion

This paper investigates the financial flexibility preference of Japanese firms during the global financial crisis period. The main results are as follows: (1) in the pre-crisis period, firms with more growth opportunity accumulate financial flexibility, (2) in the post-crisis period, firm with less growth opportunity accumulate financial flexibility, (3) in the post-crisis period, financially constrained firms accumulate financial flexibility more aggressive than the others, (4) these results are robust for an alternative measure of financial flexibility as excess cash.

Japan is known for low growth opportunity, it is not surprising that firms accumulate cash holdings and decrease net leverage when they have few investments. Previous literature on Japanese conservative debt policy argue that this negative effect of growth opportunity on financial flexibility dominants Japanese firms' debt policy, however, this paper provides another explanation for this phenomenon. At least in normal time, Japanese firms pursue financial flexibility for investment opportunity as traditional financial flexibility theory argues. The uncommon negative effect only appears after the global financial crisis, the deteriorated investment environment

triggers this negative effect of growth opportunity on financial flexibility and financial constraint deepens it.

For financially constrained firms, the experience of capital shortage during the financial crisis may force them to save aggressively in the post-crisis period, substantial financial support is expected to ease their difficulties. Furthermore, the switch of relation between growth opportunity and financial flexibility may also arise from the different financing cost that the firms faced and therefore different order between cash and debt in different periods [22], which should be further investigated.

## References

- [1] DeAngelo H, DeAngelo L. Capital Structure, Payout Policy, and Financial Flexibility. Marshall School of Business Working Paper 2007; No. FBE 02-06. Los Angeles, CA: University of Southern California.
- [2] Denis D J. Financial flexibility and corporate liquidity. *Journal of Corporate Finance* 2011; 17: 667-674.
- [3] Graham J R. How big are the tax benefits of debt. *Journal of Finance* 2000; 55(5): 1901-1941.
- [4] Bessler W, Drobetz W, Haller R, Meier I. International Zero-leverage Phenomenon. *Journal of Corporate Finance* 2013;23: 196-221.
- [5] Takami Shigeo. Factors inhibiting Japanese firms from zero leverage: financial constraints and bank relationships. *Asia-Pacific Journal of Accounting & Economics* 2016; 23(2): 161-176.
- [6] Byoun S, Xu Z. Why Do Some Firms Go Debt-Free? *Asia-Pacific Journal of Financial Studies* 2013;42(1): 1-38.
- [7] Dang V A. An Empirical Analysis of Zero-leverage Firms: Evidence from the UK. *International Review of Financial Analysis* 2013; 30: 189-202.
- [8] Minton B A, Wruck K H. Financial Conservatism: Evidence on Capital Structure from Low Leverage Firms. Working Paper, 2001. Columbus, Ohio State University.
- [9] Opler T, Pinkowitz L, Stulz R, Williamson R. The determinants and implications of corporate cash holdings. *Journal of Financial Economics* 1999; 52: 3-46.
- [10] Strebulaev I A, Yang B. The Mystery of Zero-leverage Firms. *Journal of Financial Economics* 2013; 109(1): 1-23.
- [11] Cui W. Is debt conservatism the solution to financial constraints? An empirical analysis of Japanese firms. *Applied Economics* 2020;52(23): 2526-2543.
- [12] Gamba A, Triantis A J. The value of financial flexibility. *Journal of Finance* 2008; 63(5): 2263-2296.
- [13] Gilbson M. Can bank health affect investment? Evidence from Japan. *Journal of Business* 1995; 68: 281-308.
- [14] Mian A, Sufi A. The great recession: lessons from microeconomics data. *American Economic Review* 2010; 100: 51-56.
- [15] Bliss B A, Cheng Y, Denis D J. Corporate payout, cash retention, and the supply of credit: Evidence from the 2008-2009 credit crisis. *Journal of Financial Economics* 2015; 115: 521-540.
- [16] McConnell J J, Servaes H. Equity ownership and the two faces of debt. *Journal of Financial Economics* 1995; 39: 131-157.
- [17] Almeida H, Campello M, Weisbach M S. The cash flow sensitivity of cash. *Journal of Finance* 2004, 59(4), 1777-1804.
- [18] Faulkender M, Wang R. Corporate financial policy and the value of cash. *Journal of Finance* 2006; 61: 1957-1990.
- [19] Duchin R, Ozbas O, Sensoy B A. Costly external finance, corporate investment, and the subprime mortgage credit crisis. *Journal of Financial Economics* 2010; 97: 418-435.
- [20] Iona A, Leonida L, Ventouri A. Does executive ownership lead to excess target cash? The case of UK firms. *Corporate Governance* 2012;17(5): 876-895.
- [21] Almeida H, Campello M, Weisbach M S. Corporate financial and investment policies when future financing is not frictionless. *Journal of Corporate Finance* 2011; 17: 675-693.
- [22] Acharya V V, Almeida H, Campello M. Is cash negative debt? A hedging perspective on corporate financial policies. *Journal of Financial Intermediation* 2007;16: 515-554.

# Application of Deep Reinforcement Learning Algorithm in Smart Finance

Chunhui CHEN<sup>1</sup> and Yichun ZHOU

*Guangdong University of Foreign Studies, China*

**Abstract.** Finance is not only the lifeblood of an economy, but also the lever to adjust the macro-economy. A modern economy is a market economy and essentially a developed financial economy. Based on the analysis of the problems faced by traditional finance and the overview of smart finance, this study puts forward the application of deep learning combined with reinforcement learning in smart finance to solve the problems existing in financial activities for the first time, and verifies through experiments. The model has better data and information processing ability compared with the traditional financial analysis mode. It provides higher quality decision-making information and bring more benefits. Taking a bond rating report as an example, it usually takes about 2 hours for manual in-depth analysis and carding, while it only takes about 2 minutes to interpret and refine the report by using the deep reinforcement learning model. The model has a certain reference value to solve the problems of traditional finance.

**Keywords.** Deep learning, Reinforcement learning, Deep reinforcement learning, Smart finance, Value quantity, Behavior decision-making

## 1. Introduction

With the popularity of the concept of wisdom, new concepts such as smart city, smart transportation, smart medical care, and smart finance are gradually emerging in peoples' lives. Driven by AI and big data the two core technologies, smart finance has become the inevitable direction of financial business transformation and high-quality development under the new situation. From the perspective of the demand side, the demand for financial services presents a trend of refinement and ecology. A single financial product and service has become increasingly difficult to meet the actual financial needs of customers. To this end, the financial business is moving towards great integration and collaboration. Its development mode needs to further transform from online and mobile to personalized and intelligent urgently. Through intelligent technology, customers' needs can be quickly perceived. And adaptive financial services can reach customers actively, quickly, and accurately. From the supply side, relying on intelligent technology, the financial business can be analyzed from descriptive and statistical analysis to diagnostic, predictive and decision-making analysis. It can dig out deeper financial business value. Intelligent products can upgrade and improve the existing wired and digital products. It can help to innovate, and improve the supply of financial products, improve the quality and efficiency of financial supply-side reform,

---

<sup>1</sup> Corresponding Author, Chunhui CHEN, Network and Information Center, Experimental Teaching Center, Guangdong University of Foreign Studies, Guangzhou, China. Email: 308469691@qq.com.

optimize the rational allocation of financial assets, reduce the cost of financial business, and enhance the ability of financial services to the real economy. Big data, artificial intelligence and other technologies have been widely used in the financial risk credit, investment research and many other fields of intelligent construction. With the increasing complexity of financial business, the construction of smart finance also puts forward higher requirements for the application and innovation of big data, AI, and other technologies. It is necessary to build a platform-based infrastructure for integrating data resources, computing resources and algorithm resources, to provide more powerful support for the construction of smart finance in data processing, algorithm modeling, and prediction. Aiming at smart finance, this study first analyzes the problems faced by traditional finance and puts forward the concept of smart finance. Then it introduces the deep reinforcement learning model. Finally, it puts forward the application of deep reinforcement learning in smart finance to make smart finance full of the characteristics of transparency, convenience, flexibility, timeliness, efficiency, and security. This study has a certain reference value for the application of artificial intelligence in the financial field.

## 2. The problems and causes of traditional finance

In the financial market, high-quality information plays an important role in the success of investors. With the modernization of financial transactions and information systems, massive financial data emerge in the financial market. The traditional financial model mainly uses the statistical analysis method to analyze the financial data. The traditional statistical method performs well when the data meet some specific conditions. But the financial price will be affected by various factors, the financial data often presents a non-linear relationship and has high volatility [1]. Therefore, in the actual application, the commonly used statistical analysis methods cannot process the massive financial data, and it is difficult to predict the movement trends of the financial market effectively. It also leads to the financial crisis in the traditional financial market. Therefore, activities of the traditional financial market are often faced with high risk and high cost. At present, there is more and more integration of computer technology to deal with business transactions. Through literature review, most of the methods used in financial analysis and forecasting are fundamental analysis and technical analysis. The premise of using these two methods is market inefficiency. In the economic environment characterized by highly integrated transaction processing information systems, the traditional statistical methods for fundamental analysis and technical analysis have become inefficient and ineffective. [2] The financial industry has the following performances: a) financial institutions are unable to grasp the static and dynamic project information of economic development in time; b) financial institutions are unable to grasp the static and dynamic development information of various industries in time; c) financial institutions are unable to grasp the static and dynamic business information of large and medium-sized enterprises in time; d) all financial institutions can not always be aware of the decisions made by investors for non fully rational people.

Deep reinforcement learning is a new data analysis technology. It abandons the premise that the data needs to meet specific conditions. It can analyze and forecast even when the market is effective. Through deep reinforcement learning, financial analysts can analyze business processes and related internal controls effectively. [3] They can

also analyze and evaluate all transactions recorded and determine deviations automatically and quantitatively. [4]

### 3. Overview of smart finance integrated with deep reinforcement learning

#### 3.1 The concept of smart finance integrated with deep reinforcement learning

Deep learning [5] is the general name of the deep neural network. It is composed of an input layer, hidden layer, and output layer [6]. The output layer is a single-layer structure. The hidden layer is a single-layer or multi-layer structure. Deep learning refers to self-learning according to training data without programming every problem to be solved. The goal of the deep learning model is to model the data, to analyze the deep correlation within the data and to establish a knowledge framework [7]. The model is used in prediction, classification, and feature extraction. There are three main learning models of deep learning: deep confidence network [8], convolutional neural network and recurrent neural network. [9] Reinforcement learning is a branch of machine learning. Reinforcement learning aims to achieve a goal through multi-step appropriate decision-making under a series of scenarios. Deep reinforcement learning [10] combines the perceptual analysis ability of deep learning with the decision-making ability of reinforcement learning. The most representative researcher is Google's DeepMind [11] team. This team has published two articles on deep reinforcement learning in nature, namely *Deep reinforcement learning algorithm based on video games and the AlphaGo program*. Compared with traditional finance, smart finance is a faster, more efficient, and safer financing. Integrated with deep reinforcement learning, smart finance is to use deep learning model to deal with many complex data. [12] It can analyze the internal correlation of data and establish the corresponding knowledge architecture. Through reinforcement learning, we can make decisions under different scenarios and make more accurate judgments on the consequences of decisions. [13] And we will achieve convenient, safe, high-speed, and efficient financial activities. The operation process of smart finance integrated with deep reinforcement learning is shown in Figure 1.

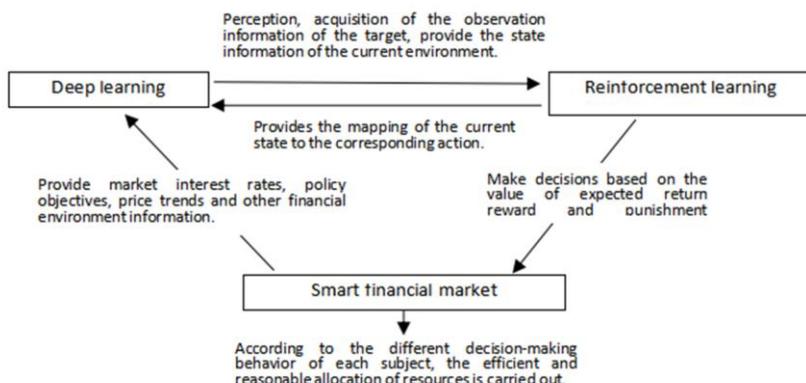


Figure 1. Operation process of smart finance integrating deep reinforcement learning

### 3.2 Basic algorithm of deep reinforcement learning

In a certain state, the value obtained by performing an action is  $Q$ ,  $\chi_t \in R^{m \times n}$  is the observation of time  $t$  ( $t=t_1, t_2, \dots, t_n$ ), symbol  $a_t \in \Lambda$  indicates the action performed under observation  $\chi_t$ .  $\Lambda$  is the set of all possible actions, the symbol  $r_i$  indicates reward (or punishment) for performing the action  $a_t$  under observation  $\chi_t$ . In addition:

$$R_t = \sum_{t'=t}^T \gamma^{(t'-t)} \cdot r_{t'} \quad (1)$$

$\gamma \in (0, 1)$  is discounted factor,  $R_t$  is the sum of all cumulative consequences from time  $t$  to  $t_n$ .

The symbol  $Q(s, a)$  is the state action-value function, the state  $s$  at time  $t$  is

$$s_t = (\chi_1, a_1, \dots, \chi_{t-1}, a_{t-1}, \chi_t) \quad (2)$$

Next, the state action-value function is optimized according to the iterative formula.

$$\begin{cases} Q_{k+1}(s_t, a_t) = Q_k(s_t, a_t) + a_k \cdot \delta_k \\ \delta_k = r_{t+1} + \gamma \cdot \max_{a' \in \Lambda} Q_k(s_t, a') - Q_k(s_t, a_t) \end{cases} \quad (3)$$

$a_k$  is the learning rate,  $s_t$  and  $a_t$  are the state and action corresponding to time  $t$ ,  $\delta_k$  and  $\gamma$  are the time difference and the discounted factor,  $a'$  express the action that state action value function can execute under the  $k$ th iteration and  $s_{t+1}$ .

The strategy of selecting the optimal executable action in a certain state is to maximize the expected value, which has the following formula.

$$\lim_{k \rightarrow \infty} Q^k = Q^*(s, a) = \mathbb{E}_{s' \sim \xi}(r + \gamma \cdot \max_{a'} Q^*(s', a') | s, a) \quad (4)$$

$\xi$  represents the environment, including all states.  $s'$  is the state after executing the action  $a$  under the state  $s$ , and  $a'$  is the possible action.

In practical application, the function approximation strategy is usually used to estimate the action-value function.

$$Q(s, a, \theta) \approx Q^*(s, a) \quad (5)$$

$\theta$  is the weight parameter to be optimized, assuming  $e_t$  plays back the experience at any time  $t$  it stores the experience at any time  $t$ .

$$e_t = (s_t, a_t, r_t, s_{t+1}) \quad (6)$$

Suppose that all actions end at the time  $t_n$ . Then the experience playback corresponds to the set  $D = [e_{t_1}, e_{t_2}, \dots, e_{t_n}]$ .

The state action value function is modified to:

$$Q(s, a, \theta) \rightarrow Q(\varphi(s), a; \theta_k)$$

And there are:

$$s_{t+1} = (s_t, a_t, \chi_t) \quad (7)$$

Experience playback is amended as follows:

$$\begin{cases} D \rightarrow \bar{D} = [\bar{e}_{t_1}, \bar{e}_{t_2}, \dots, \bar{e}_{t_n}] \\ \bar{e}_t = [\varphi(s_t), a_t, r_t, \varphi(s_{t+1})] \end{cases} \quad (8)$$

$$L_k(\theta_k) = \mathbb{E}_{\varphi(s), a \sim \rho(\cdot)} [(y_k - Q(\varphi(s), a; \theta_k))^2] \quad (9)$$

$\rho(\cdot)$  represents the behavior distribution, i.e.,  $\rho(\varphi(s), a)$  is the probability distribution of state  $\varphi(s)$  and the probability distribution of behavior  $a$ ,  $y_k$  represents the target Q (output) corresponding to the kth iteration, and has:

$$y_k = \mathbb{E}_{\varphi(s)' \sim \xi} [r + \gamma \cdot \max_{a' \in \Lambda} Q_k(\varphi(s)', a'; \theta_{k-1}) | \varphi(s), a] \quad (10)$$

The parameter  $\theta_0$  is known. After getting the target output  $y_1$ , by optimizing the objective function, we get the result  $\theta_1$ . Finally, the convergence of parameters is realized.

$$\lim_{k \rightarrow \infty} \theta_k = \theta_* \quad (11)$$

The parameters are updated by the gradient descent method.

$$\begin{aligned} \nabla_{\theta_k} L_k(\theta_k) &= \mathbb{E}_{\varphi(s), a \sim \rho(\cdot), s' \sim \xi} [r + \gamma \cdot \max_{a'} Q_k(\varphi(s)', a'; \theta_{k-1}) - Q(\varphi(s), a, \theta_k)] \cdot \\ &\quad \nabla_{\theta_k} Q(\varphi(s), a; \theta_k) \end{aligned} \quad (12)$$

Suppose that the income obtained by financial activities (such as loans, portfolio, etc.) is Q, the action of the actor is  $a$ , and the current financial environment is  $s$ .  $s$  and  $a$  are dynamic changes, and there are the following optimal income calculation functions:

$$Q = Q(s, a) \quad (13)$$

## 4. Application of deep reinforcement learning model in smart Finance

The core technology layer of applying deep reinforcement learning model to the construction of smart finance is the big data platform and AI platform. The big data platform provides the AI platform with data and computing resources needed for algorithm modeling. The AI platform can feed back the processing results of complex data to the big data platform and enrich the types of data services provided by the big data platform. The two operate in coordination to enable the construction of financial business scenarios and meet and respond to the demands of intelligent applications in different business fields quickly. The following are some application scenarios of the deep reinforcement learning model in smart finance:

### 4.1 Loan issuance

Issuing loans is the main source of income for banks. To reduce financial risks, banks need to score and screen the credit of loan applicants before making decisions on granting loans. In this process, banks need to pay costs, such as contract formulation, screening technology, customer potential default loss cost, etc. Profit is the goal of bank loans. If the bank loans can not get more than the cost of income, the bank will refuse to lend. Because the decision-making of the bank brings different incomes in different environments, we can use the above optimized income function to calculate the income of the bank in different states.

Suppose from time  $t$  to  $t_n$ , all states, the action  $a_t$  made by the bank, is making a loan (or refusing to make a loan). The amount of value obtained is  $Q$ , and the cost of the loan is  $C$ . Using the optimized income calculation function above, the target  $Q$  is solved. If  $Q > C$ , the loan will be granted (or refused). If  $Q < C$ , the loan will be refused (or granted). It is assumed that the value  $Q$  obtained by investors is obtained when their utility reaches the maximum, and is higher than that obtained by reinvestment.

#### 4.2 Default risk judgment

Here, default is relative to the behavior of the customer, such as the loan applicant. When the customer signs a loan agreement with the bank, the customer needs to pay the corresponding cost for the obtained funds within a certain period. During the contract period, customers need to use the loan funds to obtain income in order to repay the interest and principal. Under different conditions, there are differences in the ability and possibility of customers to obtain income, as well as the solvency and willingness.

Suppose from time  $t$  to  $t_n$ , all states, the action  $a_t$  made by consumers is keeping an appointment or defaulting. The value obtained for performance is  $Q$ , and the borrowing cost is  $C$ . If  $Q > C$ , the probability of customer default is small; if  $Q < C$ , the probability of customer default is greatly increased. It is assumed that the value  $Q$  obtained by investors is obtained when their utility reaches the maximum, and is higher than that obtained by reinvestment.

#### 4.3 Calculation of portfolio income

A portfolio can effectively avoid unsystematic risks. Its return is the weighted average of the returns of several securities in the portfolio. The yield of various bonds will fluctuate in different environments, so it is relatively complex to calculate the dynamic change of the portfolio yield. Often, all kinds of securities can only be combined according to the yield at a certain point in time. Using the deep reinforcement learning model, we can effectively calculate the portfolio income and maximize the income through portfolio adjustment.

Suppose that from time  $t$  to  $t_n$ , in each different state, the action  $a_t$  made by us is to choose different funds and shares for the portfolio. The value obtained is  $Q$ . Using the above optimized income calculation function, we can solve the target  $Q$  when we make different behaviors in the dynamic environment. By comparing the value quantity  $Q$ , the optimal action is selected to maximize the income. It is assumed that the value  $Q$  obtained by investors is obtained when their utility reaches the maximum, and is higher than that obtained by reinvestment.

### 5. Model validation

The above-mentioned research on the application of deep reinforcement learning in smart finance has achieved certain results. The main extracted contents are shown in Table 1.

**Table 1.** Summary of main research results

Core mechanism	Problem solving and advantages	Deficiencies	Applicable scenarios
Objective function, objective network and experience playback	Realize the prediction and decision-making under different states, and apply to the income judgment and decision in financial problems	The calculation depends on the state and can't get the specific strategy, so it is not suitable for the too complex state	It is mainly used in portfolio allocation and income calculation

Corresponding to different application goals, data assets also have rich types. Based on financial scenarios such as wealth management, investment research and risk internal control, this study mainly constructs two types of data assets: feature factor library and knowledge map. The dynamic feature factor library is based on the original data. It is combined with the business logic thinking in the financial field to build a feature index. The feature index can be effectively applied to business analysis and modeling prediction. It is also updated with the evolution of the data continuously, such as the basic information of the customer's age and city, the financial information of mortgage value, the purchase of funds in the stock type, and so on. The financial knowledge map is mainly to integrate structured, semi-structured, unstructured, and other different types of data in the financial field through a large-scale semantic network. These data can be shown in the form of graph connection to precipitate them into a structured knowledge system. Through further analysis and excavation, deeper hidden information is obtained. It makes information retrieval and query more intelligent. It also shows the relationship between different subjects more clearly, studies and judges possible risk events. In addition to the application performance, the algorithm is verified from the requirements of equipment and learning time. Table 2 shows the learning time and computer equipment of the algorithm. Table 3 shows the comparison of the time and effect of using statistical methods and the algorithm to analyze the bond investment report.

**Table 2.** Learning time and equipment of basic deep reinforcement learning algorithm --DQN algorithm

Learning time	Computer type
8 days	CPU

**Table 3.** The time and effect of statistical analysis and the deep learning algorithm in analyzing a bond investment research report

	Time required	The results form	Annualized yield
Statistical analysis	About 2 hours	Words and sentences; Weak correlation	4%-6%
Deep reinforcement learning algorithm	About 2 minutes	Atlas; Strong relevance	6%-10%

The proportion of the annual stock of debt securities assets invested by investors in the total annual debt securities investment assets is taken as the weight of different bonds. The average monthly yield of five-year treasury bonds in the current year is taken as the multiplication factor. Then, the annual weighted average long-term bond yield of each region is taken as the yield of investors' bond investment in that year. To sum up, compared with the traditional finance, in the smart finance mode of deep reinforcement learning, the actors can make more efficient decisions and obtain higher income.

## 6. Conclusion

In this study, we take the construction of smart finance as the research object. We discuss the application of deep reinforcement learning algorithm to the field of smart finance in the upsurge of artificial intelligence. Compared with the traditional statistical analysis method of finance, the application of deep reinforcement learning in smart finance can obtain valuable information from many nonlinear and complex financial data efficiently. It also helps the actors in financial activities to make more efficient decisions. Through the application of deep reinforcement learning in smart finance, the entire financial system faces lower risks and obtains higher returns. In the practice of business intelligence, financial institutions need to combine the data system, technology system and application system. It also needs to build a smooth vertical interaction architecture among the big data platform, the AI platform, and the application platform to realize the landing application of intelligent products in pricing, risk assessment, trend prediction and other scenarios. The results show that the application of deep reinforcement learning in smart finance can bring higher benefits for financial activities.

## 7. Limitations and Prospects

In the application of the model, the updating of micro information and macro information needs timely, dynamic, diversified, and multilateral information resources. The data in the database should be updated continuously to meet these requirements. At present, we can adopt the way of classifying database construction to solve the above problems, such as the existing historical database (static database) and real-time database (dynamic database). And we can adopt the data warehouse to complete the task of providing data management. It is gratifying that many data warehouse products have their own perfect data mining tools. It reduces the time and procedures for the development process of the users' strategic decision support system, but also brings the problem of data lack of scalability. Although the deep reinforcement learning algorithm applied to the smart finance has achieved good results in the simulation test platform, it has not migrated the algorithm to the real trading environment such as investment banks, commercial banks, and exchanges. In the future, with the support of big the data platform and the AI platform, we will carry out the development of data assets and the construction of algorithm model. The deep reinforcement learning model will finally realize the landing and application of intelligent scenario products, such as investment research, risk internal control and wealth management in the financial field.

## References

- [1] Labatt S. Environmental Finance: A Guide to Environmental Risk Assessment and Financial Products[J]. Transplantation, 2002, 66(8):405-9.
- [2] Tian M W, Wang L, Yan S R, et al. Research on Financial Technology Innovation and Application Based on 5G Network[J]. IEEE Access, 2019, PP (99):1-1.
- [3] B Z L A, C D H, D X L, et al. Information structures in a covering information system[J]. Information Sciences, 2020, 507:449-471.
- [4] Li, Y. (2019). Reinforcement learning applications. arXiv preprint arXiv:1908.06973.
- [5] Qiu Liwei. Application of deep reinforcement learning in video games [D]

- [6] Scott, H, Lee, et al. Chief complaint classification with recurrent neural networks. [J]. Journal of Biomedical Informatics, 2019.
- [7] Li Y, Fang S, Bai X, et al. Parallel Design of Sparse Deep Belief Network with Multi-objective Optimization[J]. Information Sciences, 2020, 533.
- [8] Charpentier, A., Elie, R., & Remlinger, C. (2021). Reinforcement learning in economics and finance. Computational Economics, 1-38.
- [9] Fu T, Zang Y, Zou Q, et al. Using Deep Learning to Identify Molecular Junction Characteristics[J]. Nano Letters, 2020, 20(5):3320-3325.
- [10] Tang J, Dong S, Cui C, et al. Sampled-Data Modeling for Wireless Power Transfer Systems[J]. IEEE Transactions on Power Electronics, 2019, PP (99):1-1.
- [11] Groman S M, Keistler C, Keip A J, et al. Orbitofrontal Circuits Control Multiple Reinforcement-Learning Processes[J]. Neuron, 2019, 103(4):734-746.e3.
- [12] Ismail A, Khalil S, Safieddine A, et al. Smart investments by smart money: Evidence from acquirers' projected synergies[J]. Journal of Corporate Finance, 2019, 56:343-363.
- [13] Saufi M S R B M, Ahmad Z A B, Leong M S, et al. Gearbox fault diagnosis using a deep learning model with limited data sample[J]. IEEE Transactions on Industrial Informatics, 2020, PP (99):1-1.

# Tax Risk Prediction of Real Estate Based on Convolutional Neural Network

Meilin YIN<sup>1</sup> and Ning LUO

*Guangdong University of Foreign Studies, China*

**Abstract.** Risk management is an important link in tax administration. From China's taxation practice, risk identification has become the weakness of tax management. With the complexity of massive data and the secrecy of modern transactions, traditional tax risk identification can no longer adapt to the development of the times. In the past, most risk researches focused on the basic machine learning stage. There are gaps in the application of deep learning in tax risk management. Based on the tax risk management indicators, this paper took the real estate industry as an example. We used convolutional neural network (CNN) to construct a tax risk prediction model. The experiment shows that a tax risk prediction model based on CNN has higher accuracy in tax risk identification and has a stronger ability to process tax data. The model has a certain reference value for tax authorities to reduce tax risk and tax loss.

**Keywords.** Tax risk prediction, convolutional neural network

## 1. Introduction

Taxation plays an important role in organizing fiscal revenue and regulating economic operations. China has implemented tax system reforms many times. The reforms have improved the structure of tax system. But tax evasion has long existed in China. With the rapid development of big data, digital economy is currently important component of China's national economy [1]. The secrecy of those transactions affects the tax administration and risk identification. Tax risk identification is facing greater challenges, due to the diversified modern transactions, the complex financial accounting methods and the huge and complicated data. In the past, most of studies focused on the basic machine learning stage. They generally used the algorithms of random forest and BP neural network. There is a gap in the application of deep learning in tax risk management [2]. This paper takes the real estate industry as the research object. We build a tax risk prediction model of real estate industry based on convolutional neural network. The model provides scientific basis and technical support for the identification and prediction of taxation risks in China. It has certain reference value for tax risk management.

This paper presented below consists of four sections. The second section analyzes the status quo and causes of tax risk management in China and puts forward the tax risk management indicators adopted in this paper. The third section will briefly discuss the proposed methods. The fourth section will experiment and analyze the tax risk

---

<sup>1</sup> Corresponding Author: Meilin Yin, Network and Information Center, Experimental Teaching Center, Guangdong University of Foreign Studies, Guangzhou, China; E-mail: 383635698@qq.com.

prediction model adopted in this research. The fifth section elaborates the significance, the insufficiency and the direction of the follow-up research.

## 2. An overview of tax risk management

### 2.1. The current status and causes of tax risk management in China

The traditional forms of tax collection emphasize manual work. Calculation, collection and risk identification depend on the experience of tax personnel [3]. With the complexity of modern economy, the drawbacks of relying on manual greatly affect the efficiency of tax administration. The authorities have realized the importance of big data in administration. After more than 30 years, the construction of tax informatization has achieved considerable results. However, there has been a great contradiction between the advanced nature of big data and the backwardness of some grass-roots tax officials. They have not fully utilized the big data [4].

We take the real estate industry as an example. The real estate industry has played an important role in economic development. It has a high degree of relevance and a complex industrial chain [5]. Meanwhile, it involves many types of taxes. The calculation of funds is also more complicated. Therefore, the tax risk investigation of the real estate industry has become a very difficult part of the administration.

### 2.2. The main tax risk management indicators in China

According to the current tax risk management indicators of China's taxation department, and screening according to the availability of data, the tax risks are mainly divided into the following categories as the output of the research model in this paper.

- A-level risk indicators, are mainly used to monitor possible false transactions. The specific classification is as follows. ①Accounts receivable are greater than operating income. ②Accounts payable are greater than operating income. ③The total cost of the period is greater than 30% of the operating income.
- B-level risk indicators, are mainly to monitor possible tax evasion. The specific classification is as follows. ①Inventory is negative. ②Inventory is greater than 30% of operating income. ③The cost of sales is greater than the operating income. ④Accounts receivable are negative. ⑤The accounts payable are negative. ⑥The prepayment is negative.
- C-level risk indicators, are mainly used to monitor possible abnormal behaviors that are not directly related to tax items. The specific classification is as follows. ①Other receivables are greater than operating income. ②Other payables are greater than operating income.
- D-level risk indicators, are mainly used to monitor possible abnormal behavior of indirect tax-related items. The specific classification is as follows. ①Non-operating income exceeds 1% of operating income. ②Non-operating expense exceeds 1% of operating income. ③Other receivables are negative. ④Other payables are negative.

### 3. An overview of Convolutional Neural Network

#### 3.1. The concept and application of convolutional neural network

Convolutional Neural Network (CNN) is a feed-forward neural network. It is a hierarchical model imitating human neural network and includes convolution calculation and deep structure. It has good performance in image recognition, classification and positioning [6-8]. In recent years, it has also been used in many fields such as financial currency and disease recognition. CNN is one of the representative algorithms of deep learning. It can be applied to diversified scenarios, and has good computational results in risk identification and target detection [9]. However, there is a gap in the application of CNN in tax risk management in China.

#### 3.2. Convolutional neural network model

Convolutional Neural Network (CNN) consists of five basic structures, which includes input layer, convolution layer, pooling layer, fully connected layer and output layer. Each layer connects the same weighted neurons and maps them to different areas of the upper layer. We can obtain a neural network structure with translation invariance. Through the feature layers (convolutional layer and pooling layer), the original features are continuously extracted and compressed. More reliable high-level features are gradually obtained from low-level features. Then the last layer is used for tasks such as classification and regression. Its structure is shown in Figure 1.

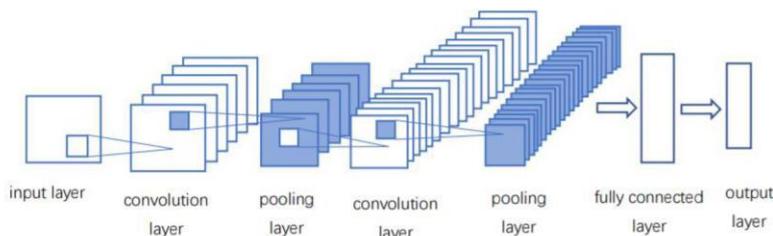


Figure 1. Structure of CNN.

The input layer is the training sample, which is the original raw data. In the convolutional layer, the input original data is divided into different regions by the convolution operation. In the actual operation process, in order to obtain higher-level features, we usually use multi-layer convolution. In simple terms, if we use the fewer convolution layers, we will get the lower feature extraction level. Multiple convolutions can make the low-level features gradually become high-level features. Pooling is mainly used to compress features. After the calculation of the fully connected layer, the result is passed to the output layer. This paper uses *Softmax* calculation to output the final classification results. The algorithm is as follows.

①The formula for calculating the  $j$  feature graph of the  $l$  layer of the convolution layer is:

$$X_j^l = f \left( \sum_{i \in M_j} (X_j^{l-1} \otimes K_{i,j}^l) + b_j^l \right) \quad (1)$$

In formula (1),  $K_{i,j}^l$  represents the convolution kernel.  $\otimes$  is the convolution operator.  $b_j^l$  represents a constant offset.  $X_j^{l-1}$  represents the feature graph of  $(l-1)$  layer, and is associated with  $X_j^l$ . ( $M_j=1, \dots, N^l$ ).

The convolution operator is an integral operation through convolution, which is used to calculate the area of the overlap region of two curves. It can be thought of as a weighted sum, replacing the pixel value of a point with a weighted average around it.

②The formula for the  $(l+1)$  pooling layer is:

$$X_j^{l+1} = p(X_j^l) \quad (2)$$

In formula (2),  $p(\cdot)$  is the down-sampling function.  $X_j^l$  represents the feature graphs of the  $l$  layer.

③The commonly used activation function is  $ReLU$ . When  $x$  is a negative number, the neuron will be 0. That is neuron necrosis. This paper uses the  $ELU$  function to avoid the death of some neurons to a certain extent. The calculation formula is:

$$y = \begin{cases} x, & x > 0 \\ \alpha(e^x - 1), & x \leq 0 \end{cases} \quad (3)$$

In formula (3),  $\alpha$  is an adjustable parameter. It controls when the negative portion of the  $ELU$  saturates. In the paper,  $\alpha=1$ .

④The formula of the fully connected layer is:

$$X_j^l = f(W_j^l \cdot X_j^{l-1} + b_j^l) \quad (4)$$

In formula (4),  $f(\cdot)$  is the  $ELU$  function.  $W_j^l$  represents the weight.  $b_j^l$  represents the offset.

⑤The formula of *Softmax* is:

$$y_i = \frac{e^{-W_i^L X^{L-1}}}{\sum_{j=1}^M e^{-W_j^L X^{L-1}}} \quad (5)$$

In formula (5),  $y_i$  is a prediction vector ( $y_i=y_1, \dots, y_M$ ).  $W_i^L$  represents the weight.  $X^{L-1}$  represents the multiple feature graphs  $X_j^l$  of the  $l$  layer.  $M$  represents the number of categories.

## 4. Tax risk prediction model for real estate based on convolutional neural network

### 4.1. Data collection and processing

This paper analyzes the subject of "Tax Risk Forecast Model for the Real Estate Industry". The data sample is the financial statement data of 142 listed companies in

the real estate industry in the four quarters of 2019. The total amount of data is 8520, and many variables are obtained, such as operating income and accounts payable. The data covers most of the listed companies in the real estate industry. There are financial data reflecting the operating conditions of taxpayers. A large number of data conform to the selected indicators. The above data sources are true and reliable, and the information is sufficient. It is expected to achieve the desired goals of the research. The selection of data in this article is carried out according to the following procedures.

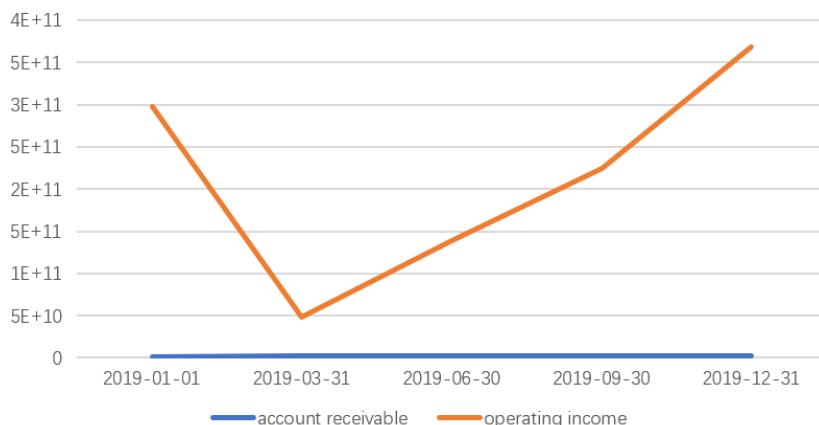
- The first step is to select the scope and content of the data according to the research and indicator requirements.
- The second step is to do a preliminary screening of the data. We will exclude data that are of little significance to the study.
- The third step is to eliminate missing value data.
- The fourth step is to classify and aggregate the scattered data items.

There are too many actual participation variables, which are limited by space and will not be listed one by one. The data comes from the CSMAR database. We took the data on December 31, 2019 as an example to screen and integrate the data, as shown in Table 1.

**Table 1.** Real estate listed companies fourth quarter financial data table.

Stock code	000002	000006	000011	Other : 123
Operating income	3.67894E+11	3731330140	3961669942	...
Operating costs	2.3455E+11	2044318167	1433615885	...
Sales expense	9044496840	51736223.03	111553952.5	...
Non-operating income	714732128.7	2516753.07	23732348.28	...
Non-operating expense	788578652.7	8100823.59	4793503.85	...
...	...	...	...	...

According to this paper adopted by the tax risk level classification method, we will have visual chart data processing. Because there are too many data, we will not enumerate them one by one. We take the relationship between 2019 accounts receivable and operating revenue with the stock code of 000002 as an example. It is shown in Figure 2.



**Figure 2.** The relationship between the 000002's accounts receivable and operating income in 2019.

#### 4.2. The model based on convolutional neural network

The obtained data is recorded as  $X$ , which is used as the input value of the model. We take the indicators, involved in the financial statement data of each enterprise, as the input of the model.

The tax risk prediction model of the real estate industry based on CNN is set as follows. Its structure is shown in Figure 3.

- Input layer: The indicators involved in the financial statement data of each enterprise are used as the input of the model.
- Convolutional layer: We use a  $5 \times 5$  convolution kernel to obtain the corresponding feature map. Then we calculate it through the ELU function and add the deviation. Finally, we use it as the value of the neuron in the C1 layer.
- Pooling layer: Filter is set to  $2 \times 2$ . Then the corresponding feature map will be output. Among them, the scaling factor is 2 to control the compression speed.
- Output layer: Since we use tax risk management indicators to classify tax risk, the final output classification uses the *Softmax* function to achieve accurate tax risk classification. For a given test sample  $X$ , there are four possible outputs. If we input a vector element, it will output the probability value of [0-1]. This probability value is the probability of each tax risk classification.

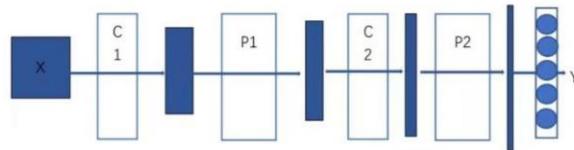
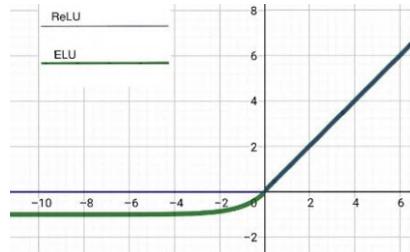


Figure 3. The model structure of tax risk prediction based on CNN.

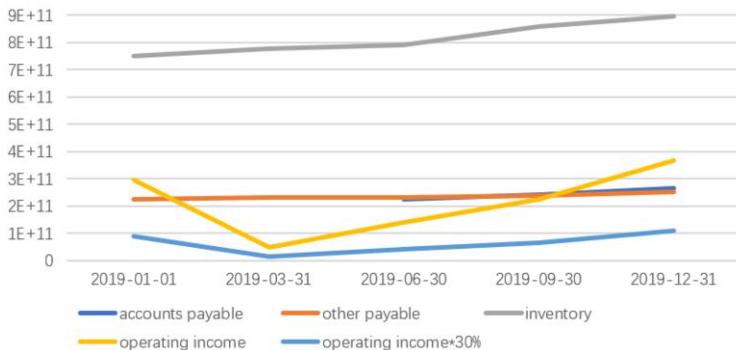
#### 5. Experimental results and analysis of tax risk prediction model based on convolutional neural network

We adopt the *ELU* function in the paper. Although the calculation amount is larger than that of the *ReLU* function, it can avoid neuron necrosis to a certain extent. When the value is negative, *ELU* has the characteristic of soft saturation. It improves the robustness to input changes. Moreover, since the output value of *ReLU* has no negative value, the mean value of the output is greater than zero. It makes the activation unit of the next layer have bias shift. The average value of *ELU* output is close to zero, which reduces the computational complexity. Therefore, compared with the traditional *ReLU* function algorithm, the convergence speed of *ELU* function is greatly improved. The function comparison is shown in Figure 4.



**Figure 4.** Functional image contrast between ReLU and ELU.

In this paper, we build a tax risk prediction model based on CNN. 80% of the data is set as the train set, while 20% of the data is used as the test set. Of the final 7260 data sets, 6000 are training sets and the remaining 1260 are test sets. In 1260 test sets, there are 1230 data consistent with the actual results, with an accuracy of 97.96%. We take the listed company with the stock code of 000002 as an example. The output result shows that the possible tax risk levels are A, B and C. Among them, the possibility with higher probability is B. We conduct data analysis on this company, as shown in Figure 5. In 2019, the company's annual inventory is more than 30% of the operating income, so there may be B-level tax risk. In the second and third quarters, accounts payable are greater than operating income, so there may be A-level tax risk. Other payables are greater than operating income, so there may be a C-level tax risk. The predicted results are consistent with the actual results.



**Figure 5.** The relationship between the variables of the 000002.

For the model, this experiment uses four evaluation indicators: accuracy, detection accuracy, recall, and F1 value to evaluate the effect of the model. The predicted results are shown in Table 2.

As shown in Table 2, the use of CNN to build a model for tax risk prediction has a high accuracy rate of up to 97.35%. It also has high overall detection and good generalization performance. This result shows that CNN has a good performance in predicting tax risks. To a certain extent, it helps the authorities to manage the tax compliance of enterprises.

**Table 2.** Test results of CNN.

	Train set	Test set
Accuracy	0.9856	0.9735
Detection	0.9825	0.9751
Recall	0.9801	0.9706
F1 value	0.9865	0.9743

## 6. Conclusion

This paper takes real estate industry as the research object, and uses convolutional neural network algorithms to build tax risk prediction models. The experiment proves that CNN can be effectively used for tax risk prediction. Compared with traditional models, CNN can effectively and accurately identify the tax risks. The model can prevent and reduce tax risks, and provides an important reference for authorities. It can also help to improve corporate tax compliance of real estate companies and other industries. It has important theoretical and practical significance for preventing and reducing tax losses, and increasing the country's fiscal revenue.

The research of this paper has two deficiencies. First, we only predict the tax risk of the real estate industry of listed companies, with limited sample data. Second, only the deep learning method of CNN is used for model prediction, and there is a lack of more research and comparison of other deep learning models. Based on the above shortcomings, the research work of this article can be expanded from the following aspects. The first one is to improve and enrich machine learning algorithms. For example, deep learning algorithms such as Recurrent Neural Network (RNN) can be included in the comparative experiment. The second one is to expand the types and scope of data, so that the research is not only aimed at listed companies, and it can be applied to a wider range of different industries. Third, other researchers use various tax risk indicators, and we can increase our research on indicator selection in subsequent studies. In future studies, improvements and further studies will be made on the above aspects.

## References

- [1] Xingyi Song, Yongsheng Song. Path selection of tax risk management under big data environment. *Taxation Research*. 2020,(03).
- [2] Weiling Wang, Jing Wang. Research on the Development Trend and Promotion Policy of my country's Digital Economy. *Economic Review Journal*. 2019 (01), 69-75.
- [3] Zhijuan Lin. Study on the Tax Management Path of Colleges from the Perspective of Internal Control. 2018 International Conference on Economics,Finance,Business, and Development (ICEFBD 2018). 2018 Oct.
- [4] Yuanzhao Gao, Xingyuan Chen, Xuehui Du. A Big Data Provenance Model for Data Security Supervision Based on PROV-DM Model. *IEEE Access*. 2020 Feb. PP(99):1-1.
- [5] Cunyi Sun. Tax risk management from the perspective of big data. *Taxation Research*. 2019 Jul. 107-111.
- [6] Fengmei Cui. Deployment and integration of smart sensors with IoT devices detecting fire disasters in huge forest environment. *Computer Communications*. 150 (2020), 818-827.
- [7] M. Anbarasan, BalaAnand Muthu. Detection of flood disaster system based on IoT, big data and convolutional deep neural network. *Computer Communications*. 150 (2020), 150-157.
- [8] S. Jothi Shri, S. Jothilakshmi. Crowd Video Event Classification using Convolutional Neural Network. *Computer Communications*. 147 (2019) 35-39.
- [9] Min Chen, Yixue Hao, Kai Hwang, Lu Wang. Disease Prediction by Machine Learning Over Big Data From Healthcare Communities. *IEEE Access*. 2017 Apr. PP(99):1-1.

# Krylov Subspace Methods for Big Data Analysis of Large Computational Electromagnetics Applications

Bruno CARPENTIERI <sup>a,1</sup>

<sup>a</sup> Faculty of Computer Science, Free University of Bozen-Bolzano, 39100 Bolzano, Italy

**Abstract.** In this paper we present some computational techniques based on the class of preconditioned Krylov subspace methods that enable us to carry out large-scale, big data simulations of Computational Electromagnetics applications modeled using integral equations. This analysis requires the solution of large linear systems that cannot be afforded by conventional direct methods (based on variants of the Gaussian elimination algorithm) due to their high memory costs. We show that, thanks to the development of efficient Krylov methods and suitable preconditioning techniques, nowadays the solution of realistic electromagnetic problems that involve tens of million (and sometimes even more) unknowns, has become feasible. However, the choice of the best class of methods for the selected computer hardware and the given geometry remains an open problem that requires further analysis.

**Keywords.** Computational Electromagnetics, Boundary Element Method, Surface Integral Equations, Krylov Subspace methods, Preconditioning, Big Data.

## 1. Introduction

Our society is characterized by an unprecedented ability to generate large volume of data that can be analysed by advanced computer algorithms to produce new knowledge, and use such information to simulate and reproduce the complex behaviour of real-world systems. The “big data” paradigm in science is driving the development of novel scientific methodologies that are opening up new frontiers for advanced discoveries in many fields of science; see e.g. [1,2,3,4,5,6] for some examples of big data simulations in computational science. In this paper, we discuss a case study in Computational Electromagnetics.

Many wave propagation phenomena in science are formulated mathematically in terms of integral equation models that are defined on the boundary of the pertinent computational domain. One of the most popular integral equation models for electromagnetic (EM) scattering applications is known as Electric Field Integral Equation (EFIE). It solves the integral equation below for the surface current  $\vec{j}$

---

<sup>1</sup>Corresponding Author: Bruno Carpentieri; E-mail: bcarpentieri@gmail.com.

$$\int_{\Gamma} \int_{\Gamma} G(|y-x|) \left( \vec{j}(x) \cdot \vec{j}'(y) - \frac{1}{k^2} \operatorname{div}_{\Gamma} \vec{j}(x) \cdot \operatorname{div}_{\Gamma} \vec{j}'(y) \right) dx dy = \frac{i}{k Z_0} \int_{\Gamma} \vec{E}_{inc}(x) \cdot \vec{j}'(x) dx,$$

where  $\vec{j}'$ 's denote suitable tangential test functions. By symbol  $\operatorname{div} \vec{j}(x)$  we indicate the divergence operator, and  $G(|y-x|) = \frac{e^{ik|y-x|}}{4\pi|y-x|}$  is the Green's function for scattering problems. The other symbols are defined as follows:  $\Gamma$  is the boundary of the domain of interest, by  $Z_0 = \sqrt{\mu_0/\epsilon_0}$  the impedance of free space, and finally by  $k$ ,  $\epsilon$  and  $\mu$  the wave number, the electric permittivity and the magnetic permeability of the medium, respectively. This model is applicable to generic objects, including those that have cavities and/or disconnected parts.

The numerical discretization of EFIE by the Method of Moments (MoM) [7] using a mesh with  $n$  edges yields an  $n$ -dimensional linear system

$$Ax = b. \quad (1)$$

In Eq. (1), the unknowns in vector  $x$  represent the vectorial flux of the surface electric current across the edges of the underlying mesh,  $A$  is a dense, complex, symmetric non-Hermitian matrix containing the contributions to the EFIE singular integrals, while the right-hand side vector  $b$  depends on the characteristics of the illuminating radiation such as the incidence angle. Scattering analysis may demand very large computer resources and highly efficient numerical algorithms. For example, accurate modeling of a perfectly conducting sphere of diameter of 1,800 wavelengths yields systems with more than 3 billion equations, whose storage requires 144,000 petabyte of data [8]. Systems of this size are not affordable using variants of the conventional Gaussian elimination algorithm. They can be solved only using matrix-free computational techniques such as the class of iterative Krylov subspace methods, that can overcome the memory bottlenecks of Gaussian elimination since they are based on matrix-vector and vector-vector operations.

The Generalized Minimum Residual (GMRES) method introduced by Saad and Schultz in [9] is a very popular Krylov subspace algorithms. After  $k$  iterations it computes the approximate solution of Eq. (1) that minimizes the 2-norm of the residual over the Krylov space  $K_k(A, r_0) = \operatorname{span} \{r_0, Ar_0, \dots, A^k r_0\}$  at the cost of  $\mathcal{O}(nk)$  arithmetic operations and storage units. On the other hand, *nonoptimal* Krylov methods are developed upon three-term vector recurrences and have  $\mathcal{O}(n)$  complexity in both time and space. The principal developments of nonoptimal methods include the Conjugate Gradients Squared (CGS) method by Sonneveld, the Quasi-Minimal Residual (QMR) and the Transpose-Free Quasi-Minimal Residual methods by Freund and Nachtigal, the Biconjugate Gradient STABilized (BiCGSTAB) method by van der Vorst, and others. See [10] for an overview of Krylov subspace methods. Recently, a new class of Krylov algorithms built upon the Lanczos  $A$ -orthonormalization procedure has shown competitive convergence rates for solving surface integral equations [11].

In many EM applications, including scattering analysis, microwave and millimeter-wave circuits design, antenna array simulations and others, the pertinent linear systems to solve have all the same coefficient matrix  $A$  and a set of different right-hand side vectors. In this circumstance, Eq. (1) writes in the form

$$AX = B, \quad (2)$$

where  $B = [b_1, b_2, \dots, b_p] \in \mathbb{C}^{n \times p}$  is the matrix of the, say  $p$ , right-hand side vectors  $b_i$ ,  $i = 1, 2, \dots, p$  and  $X \in \mathbb{C}^{n \times p}$  is the solution matrix to be computed. Block variants of Krylov subspace methods are significantly more robust than standard Krylov algorithms for solving Eq. (2) as they typically use much larger search spaces and a block implementation of the matrix-vector product operations that can better exploit the sophisticated memory hierarchy of modern computers [10].

## 2. Preconditioning boundary integral equations

The convergence of Krylov subspace methods is often slow and needs to be accelerated by a technique called *preconditioning*, which transforms the initial system  $Ax = b$  into an equivalent system that has more favourable eigenvalues distribution, *i.e.* the vast majority of its eigenvalues are grouped close to point one of the spectrum. The new transformed system writes as  $M^{-1}Ax = M^{-1}b$  if the *preconditioner matrix*  $M$  is applied from the left, or  $AM^{-1}y = b$  (here  $x = M^{-1}y$ ) if it is applied from the right. Preconditioning is necessary on EFIE as the number of Krylov iterations tends to increase as  $\mathcal{O}(n^{0.5})$  when the number of unknowns,  $n$ , is related to the wavenumber. An effective preconditioner for solving surface integral equations should be cheap to compute and easy to combine with the data structure of fast integral equations solvers, e.g., the Multilevel Fast Multipole Algorithm (MLFMA) [12],  $\mathcal{H}$ -matrices [13], wavelet techniques [14], panel clustering [15] and similar approaches, so that it can maintain overall  $\mathcal{O}(n \log n)$  complexity. Another important requirement is that it should scale satisfactorily with the frequency of the problem and the number of processors, yielding robust convergence across a wide range of geometries and physical parameters. These requirements often contradict with each other. However, when an effective preconditioner is available, the selection of the Krylov algorithm to use is much less critical.

The need to solve large dense linear systems in big data EM simulations has led to the production of many efficient methods rather than to the specialization in one specific technology. For memory concerns, many efficient preconditioners for surface integral equations in EM are constructed from a sparse matrix  $S$  that approximates  $A$  and is much easier to invert than  $A$ . Initially, the boundary element matrix  $A$  is decomposed in the form

$$A = A_{near} + A_{far},$$

where  $A_{near}$  is the block diagonal and near-diagonal part of  $A$  coming from the interactions of nearby basis functions in the mesh, while  $A_{far}$  is the far-field part of  $A$  associated with interactions of distant basis functions. Then, it is natural to define  $S = A_{diag} + A_{near}$ . In the next sections, we identify some of the most important classes of preconditioning methods for Krylov subspace solvers constructed from such approximation  $S$ .

### 2.1. Incomplete LU (ILU) factorization methods

Incomplete LU factorization preconditioners decompose approximately matrix  $S$  as  $M = \tilde{L}\tilde{U} \approx S$ , where the factors  $\tilde{L}$ ,  $\tilde{U}$  are obtained by applying an incomplete Gaussian elimination procedure to  $S$ . Due to the indefiniteness of  $A$ , on EFIE it is likely to

encounter small pivots during the factorization, and to compute ill-conditioned triangular solves [16]. The problem of ill-conditioning can be remedied by reordering the entries of  $S$  before the factorization, by shifting the diagonal entries of  $S$ , or by using numerical pivoting. A recently developed family of multilevel incomplete LU factorization methods has yielded robust performance on this problems class (see e.g. some results in [17]). Incomplete LU factorization algorithms are not inherently parallel. However, good scalability can be obtained using domain decomposition techniques at the cost of moderate computational overhead.

## 2.2. Sparse approximate inverse methods

Considerable attention in the last years in EM have received computational methods that explicitly approximate and store the inverse of boundary integral equation matrices [18,19]. The approximate inverse matrix  $M \approx S^{-1}$  is used as a preconditioner for Krylov methods. This approach is especially interesting for parallelism because applying  $M$  at every iteration simply requires one (or sometimes two) sparse M-V products that are easier to implement efficiently than conventional triangular systems solves on modern distributed memory multiprocessor computers and graphics processing units (GPUs). Admittedly, an efficient parallel implementation of these methods based on the Message Passing Interface and Open Multi-Processing paradigms with optimized parameter setting, runtime environment, load balance and minimised data movement on the given problem and dataset is a challenge in its own right [18,20]. Due to the highly localized coupling of the edges in the underlying mesh, boundary element matrices exhibit a good deal of regularity. As a consequence of this property,  $A$  can be effectively approximated by a very sparse matrix. Figure 1 shows that 1) the distribution of the large matrix entries in  $A$  and  $A^{-1}$  can be very similar because of the exponential decay of the Green's function, 2) a sparse matrix can approximate  $A^{-1}$  very effectively, and 3) the pattern of the sparsified matrix  $A$  can be a good choice for the nonzero structure of the approximate inverse.

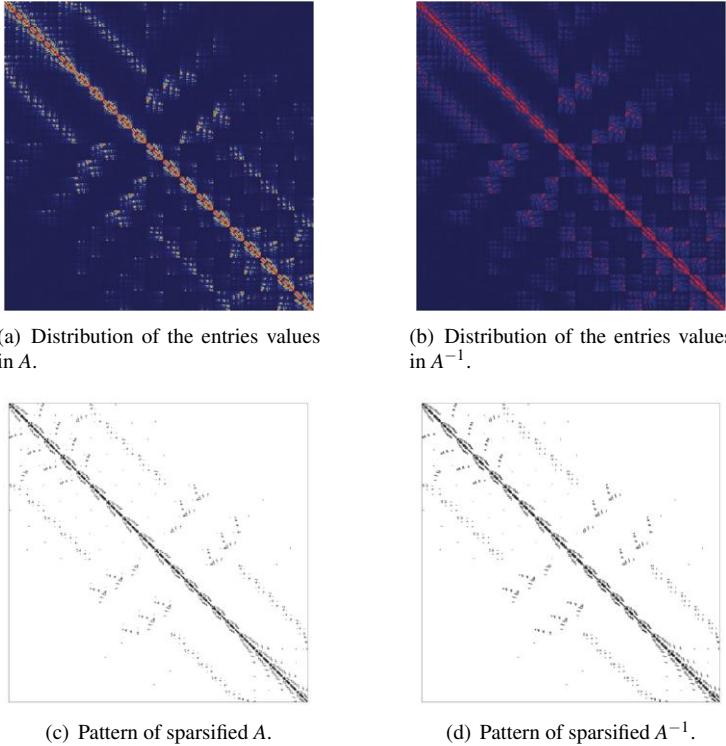
The sparse approximate inverse known as SPAI computes the numerical values of the entries of the preconditioner  $M$  by minimizing the Frobenius-norm of the matrix  $\|I - SM\|_F$  (in the case of right preconditioning), or  $\|I - MS\|_F$  (in the case of left preconditioning). The computation reduces to solving  $n$  separate linear least-squares problems, one per column or row of  $M$  depending whether the preconditioner is computed from the right or from the left, according to the following equation

$$\|I - SM\|_F^2 = \sum_{j=1}^n \|e_j - Sm_{\bullet j}\|_2^2, \quad (3)$$

where we denote by  $e_j$  the  $j$ th canonical unit vector and  $m_{\bullet j}$  is the  $j$ th column of  $M$ . For right preconditioning, an analogous relation to Eq. (3), that is

$$\|I - MS\|_F^2 = \|I - S^T M^T\|_F^2 = \sum_{j=1}^n \|e_j - S^T m_{j\bullet}\|_2^2$$

holds, where  $m_{j\bullet}$  is the  $j$ th row of  $M$ . Successful experiments with approximate inverse methods are reported for the solution of both surface and surface-volume integral equations [18,19,21].

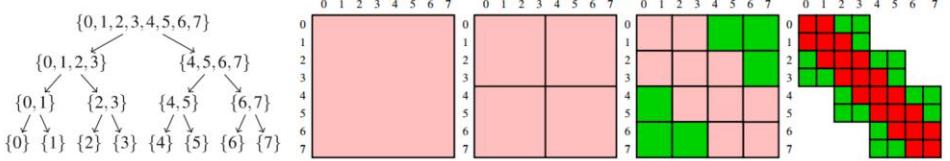


**Figure 1.** Above: typical pattern of the large entries of the coefficient matrix  $A$  (on the left) and of its inverse  $A^{-1}$  (on the right) for boundary element matrices. We depict large to small entries using different colors, from red to green, yellow and blue. The model problem is a sphere. Below: nonzero structure of  $A$  (left) and of  $A^{-1}$  (right) after thresholding all the entries having relative magnitude less than  $5.0 \times 10^{-2}$ .

### 2.3. Multilevel methods

Due to the sparsity of the local approximation  $S$  used to compute the preconditioner  $M$ , some strategies need to be developed to make the preconditioner more robust on large problems. In the inner-outer two-levels iterative method proposed in [18], for the preconditioning operation we carry out some iterations of an inner Krylov method, to balance the locality of the preconditioner with the use of the MLFMA matrix arising from the discretization. The effectiveness of this approach depends on three main ingredients: 1) the outer solver must accommodate variable preconditioners (e.g., FGMRES [22] and GMRES $\star$  [23, p. 91] are two possibilities), 2) the inner solver needs to be preconditioned to ensure a significant reduction of the inner residuals in a few iterations, 3) the matrix-vector operations in the inner solver can be less accurate as they are used for the preconditioning operation. Another attempt to improve conventional preconditioners for boundary integral equations “removes” the negative effects that very small eigenvalues of the preconditioned matrix can have on the convergence [24].

By using a three-levels inner-outer schemes, we solved a rectangular cavity problem discretized with 12,697,120 unknowns in 58 minutes. The construction of the preconditioner took 30 minutes on 36 processors, and demanded 96GB memory in total. On a computer with 120 cores of a computer cluster equipped with 6-core processors,



**Figure 2.** Left. A cluster tree in the  $\mathcal{H}$ -matrix approach. Right. Four levels block cluster tree: nodes to be refined are coloured light red, admissible leaves are green, inadmissible ones are red.

2.66GHz clock rate, 10 nodes with 48GB of memory each, a 40Gb/s Infiniband network, we solved a scattering problem from a Cobra geometry modeled with 21,682,980 unknowns in 38 minutes of CPU elapsed time, and a tank with 10,768,581 unknowns in only 12 minutes.

#### 2.4. $\mathcal{H}$ -matrix-based solvers

A more recent approach to build low-complexity data-sparse matrix solvers for EFIE linear systems is based upon the hierarchical  $\mathcal{H}$ -matrix representation of  $A$ , which replaces dense blocks of  $A$  by low-rank approximants computed without any knowledge of the underlying kernel. In the original  $\mathcal{H}$ -matrix formulation [13], the blocks that are admitted a low-rank representation satisfy the *admissibility criterion*

$$\min(diam(B_s), diam(B_t)) \leq \eta \cdot dist(B_s, B_t) \quad (4)$$

for some  $0 < \eta$ . In (4),  $B_s$  and  $B_t$  are rectangular boxes surrounding two clusters  $s$  and  $t$  of nodes in the mesh, whereas the distances ( $dist$ ) between  $s$  and  $t$  and the diameters ( $diam$ ) of these clusters are computed in terms of the Euclidean norm from the center of gravity of the box. Upon recursive partitioning of the bounding box into smaller boxes until the admissibility condition (4) is not satisfied, a cluster tree data structure is obtained, similar to the left Figure 2. Then, a hierarchical  $\mathcal{H}$ -matrix block partitioning of  $A$  is produced by associating the matrix block  $(A_{ij})_{i \in s, j \in t}$  to the cartesian product  $s \times t$ , as in the right Figure 2. It has been shown that by means of fast low-rank ( $k$ ) compression of the admissible blocks, an almost optimal  $\mathcal{O}(nk\log n)$  memory complexity and  $\mathcal{O}(nk^2 \log^2 n)$  arithmetic costs can be achieved for the matrix vector multiply, factorization and inverse operations using  $\mathcal{H}$ -matrices [25].

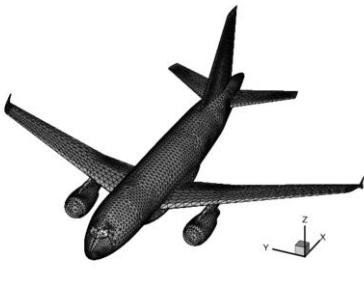
The so-called hierarchical  $\mathcal{H}^2$ -matrix representation can significantly improve the efficiency of  $\mathcal{H}$ -matrices by computing nested low rank factorizations of entire collections of blocks instead of decomposing each admissible block separately [26]. An admissible block  $A_{st}$  associated to the cluster of nodes  $s$  and  $t$  satisfies the admissibility condition

$$\max(diam(B_s), diam(B_t)) \leq \eta \cdot dist(B_s, B_t)$$

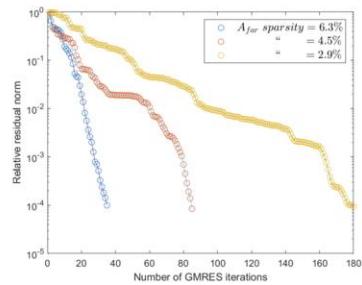
which replaces condition (4) in the theory of  $\mathcal{H}^2$ -matrices, and is represented mathematically by the rank- $k$  factorization

$$\tilde{A}_{st} = V_s S_k V_t^H,$$

where  $V_s$  of size  $\#s \times k$  and  $V_t$  of size  $k \times \#t$  (here symbol  $\#$  denotes the set cardinality) are called *cluster bases* of clusters  $s$  and  $t$ , while  $S_k$  of size  $k \times k$  is called *coupling matrix*. The  $\mathcal{H}^2$ -matrices variant of  $\mathcal{H}$ -matrices achieves  $\mathcal{O}(nk)$  storage complexity for a  $n \times n$  boundary element matrix, instead of the  $\mathcal{O}(nk \log n)$  memory complexity of standard  $\mathcal{H}$ -matrices, at the cost of only slightly larger error [26]. In Figure 3 we show some convergence results obtained by the author for solving a radar cross-section calculation problem on an Airbus aircraft prototype mesh discretized with  $n = 23676$  nodes using an  $\mathcal{H}^2$ -matrices-based solver. We can see that fast convergence can be achieved on this difficult problem using low to moderate sparsity levels for the matrix  $A_{near}$  used to construct the preconditioner.



(a) Aircraft prototype mesh (23676 dofs).  
Courtesy of EADS-CCR Toulouse.



(b) Convergence histories on the Airbus aircraft prototype mesh.

**Figure 3.** GMRES convergence histories with  $\mathcal{H}$ -matrix-based solvers on the Airbus aircraft problem.

### 3. Conclusions

In this paper we have discussed large-scale simulations of Computational Electromagnetics applications modeled using integral equations. We have shown that their rigorous numerical solution requires to process large volumes of data, and thus it is highly demanding of innovative algorithms. An essential ingredient of the big data analysis is the numerical solution of high-dimensional linear systems that cannot be solved by the standard Gaussian elimination method. We have presented low-complexity iterative solution techniques based on Krylov subspace methods, incomplete factorizations, sparse approximate inverses, multilevel schemes and  $\mathcal{H}$ -matrix-based solvers for this class of problems. An iterative method can solve an  $n \times n$  dense linear system arising from the boundary element discretization of integral equations in  $\mathcal{O}(n \log n)$  arithmetic operations, that is a dramatic improvement compared to the  $\mathcal{O}(n^3)$  work required by a direct method. Effective preconditioners are mandatory to use to decrease the total number of iterations from  $\mathcal{O}(n)$  to  $\mathcal{O}(1)$ . By using Krylov subspace methods and robust preconditioners, big data analysis of extremely large Computational Electromagnetics applications with tens of million unknowns, and even more, is becoming feasible. Some of these techniques can be used in other electromagnetic simulations, e.g. in fusion energy research [27]. However, the choice of the best class

of methods for the selected computer architecture and given geometry remains an open problem that requires further analysis.

## Acknowledgements

The author is a member of the *Gruppo Nazionale per il Calcolo Scientifico* (GNCS) of the Istituto Nazionale di Alta Matematica (INdAM) and this work was partially supported by INdAM-GNCS under Progetti di Ricerca 2020. Support from the Research Südtirol/Alto Adige 2019 grant (Provincia autonoma di Bolzano/Alto Adige – Ripartizione Innovazione, Ricerca, Università e Musei, contratto no. 19/34) is gratefully acknowledged.

## References

- [1] J. N. Kutz. *Data-driven modeling & scientific computation: methods for complex systems & big data*. Oxford University Press, 2013.
- [2] S. Velamparibil, W. C. Chew, and J. Song. 10 million unknowns: is it that big? [computational electromagnetics]. *IEEE Antennas and Propagation Magazine*, 45(2):43–58, 2003.
- [3] H. Hangal, M. Refan, C. Jubayer, D. Parvu, and R. Kilpatrick. Big data from big experiments. the windeee dome. In *Whither Turbulence and Big Data in the 21st Century?*, pages 215–230. Springer, 2017.
- [4] J. H. Faghmous and V. Kumar. A big data guide to understanding climate change: The case for theory-guided data science. *Big data*, 2(3):155–163, 2014.
- [5] Md Altaf-Ul-Amin, F. M. Afendi, S. K. Kiboi, and S. Kanaya. Systems biology in the context of big data and networks. *BioMed research international*, 2014, 2014.
- [6] S. Pal, S. Mondal, G. Das, S. Khatua, and Z. Ghosh. Big data in biology: The hope and present-day challenges in it. *Gene Reports*, 21:100869, 2020.
- [7] W.C. Gibson. *The method of moments in electromagnetics*. Boca Raton, FL: Chapman & Hall/CRC. xv, 272 p. \$ 119.95 , 2008.
- [8] B. Michiels, J. Fostier, I. Bogaert, and D. De Zutter. Full-wave simulations of electromagnetic scattering problems with billions of unknowns. 63(2), 2015.
- [9] Y. Saad and M.H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Scientific and Statistical Computing*, 7:856–869, 1986.
- [10] Y. Saad. *Iterative methods for sparse linear systems*. SIAM, Philadelphia, 2nd edition, 2003.
- [11] Y.-F. Jing, B. Carpentieri, and T.-Z. Huang. Experiments with Lanczos biconjugate A-orthonormalization methods for MoM discretizations of Maxwell’s equations. *Progress In Electromagnetics Research*, 99, pages 427–451, 2009.
- [12] L. Greengard and V. Rokhlin. A fast algorithm for particle simulations. *Journal of Computational Physics*, 73:325–348, 1987.
- [13] W. Hackbusch. A sparse matrix arithmetic based on  $\mathcal{H}$ -matrices. *Computing*, 62(2):89–108, 1999.
- [14] B.K. Alpert, G. Beylkin, R. Coifman, and V. Rokhlin. Wavelet-like bases for the fast solution of second-kind integral equations. *SIAM J. Scientific and Statistical Computing*, 14:159–184, 1993.
- [15] W. Hackbusch and Z.P. Nowak. On the fast matrix multiplication in the boundary element method by panel clustering. *Numerische Mathematik*, 54(4):463–491, 1989.
- [16] B. Carpentieri, I.S. Duff, L. Giraud, and M. Magolu monga Made. Sparse symmetric preconditioners for dense linear systems in electromagnetism. *Numerical Linear Algebra with Applications*, 11(8-9):753–771, 2004.
- [17] B. Carpentieri and M. Bollhöfer. Symmetric inverse-based multilevel ILU preconditioning for solving dense complex non-hermitian systems in electromagnetics. *Progress In Electromagnetics Research*, 128:55–74, 2012.
- [18] B. Carpentieri, I.S. Duff, L. Giraud, and G. Sylvand. Combining fast multipole techniques and an approximate inverse preconditioner for large electromagnetism calculations. *SIAM J. Scientific Computing*, 27(3):774–792, 2005.

- [19] T. Malas and L. Gürel. Accelerating the multilevel fast multipole algorithm with the sparse-approximate-inverse (SAI) preconditioning. *SIAM Journal on Scientific Computing*, 31(3):1968–1984, 2009.
- [20] C. Delgado, E. García, Á. Somolinos, and M.F. Cátedra. Hybrid parallelisation scheme for the application of distributed near-field sparse approximate inverse preconditioners on high-performance computing clusters. *IET Microwaves, Antennas & Propagation*, 14(4):320–328, 2020.
- [21] J. Lee, C.-C. Lu, and J. Zhang. Sparse inverse preconditioning of multilevel fast multipole algorithm for hybrid integral equations in electromagnetics. *IEEE Transactions on Antennas and Propagation*, 52(9):2277–2287, 2004.
- [22] Y. Saad. A flexible inner-outer preconditioned GMRES algorithm. *SIAM J. Scientific and Statistical Computing*, 14:461–469, 1993.
- [23] H.A. Van der Vorst. *Iterative Krylov Methods for Large Linear Systems*. Cambridge University Press, Cambridge, UK, 2003.
- [24] B. Carpentieri. A matrix-free two-grid preconditioner for boundary integral equations in electromagnetism. *Computing*, 77(3):275–296, 2006.
- [25] M. Bebendorf. *Hierarchical Matrices*. Springer Berlin Heidelberg, 2008.
- [26] S. Börm. *Efficient Numerical Methods for Non-local Operators*. European Mathematical Society Publishing House, December 2010.
- [27] M. Bonotto, P. Bettini, and A. Cenedese. Model-order reduction of large-scale state-space models in fusion machines via Krylov methods. *IEEE Transactions on Magnetics*, 53(6):1–4, 2017.

# Industrial Economic Cooperation Between China and Nordic Countries Under the Double Circulation Pattern: Basis and Prospect

Nianlin Zhou<sup>1</sup> Manyuan Jiang and Weixiao Lai

*South China Business College, Guangdong University of Foreign Studies, Guangzhou  
510545, P.R. China*

**Abstract.** This article takes the formula of trade complementarity index to calculate the degree of trade complementarity of various industries between China and Nordic countries based on the data of import and export from 2013 to 2018. The results indicate that the trade complementarity between China and Nordic countries has three major characteristics, that is, intensive concentration of specific complementary categories, structurally complementary in bilateral export and import and complementarity based on respective comparative advantages, and clarify the dual complementarities both in inter-industry trade and intra-industry trade. Our findings show that the basis of bilateral cooperation is stable due to strong complementarity each other. In addition, expanding the volume of trade complementary goods and cooperation are not only to promote the development of dual circulation in China, but also to reduce the segmentation costs in the processes of production for both sides.

**Key words.** Industrial economic cooperation; trade complementarity index; double circulation pattern

## 1. Introduction

The core of establishing the new development pattern of double circulation at home and abroad is to stimulating mutual enhancement and to achieve more strong and sustainable development through unleashing the potentiality of domestic demand and enhancing resource utilization between Chinese and international markets [1].

The Nordic countries are the important parts of “the belt and road” and have a vital bearing on the interests of China. The development and utilization of Arctic resources will influence on international trade, the pattern of energy supply, and Chinese economy whom is the largest trading and energy consuming country in the world [2].

---

<sup>1</sup> Corresponding Author: Nianlin Zhou, South China Business College, Guangdong University of Foreign Studies, Guangzhou, China; Email: gweduyjzx@163.com. This research was funded by the Education Bureau of Guangdong Province of China (Grant No. 2018GXJK272) and the research center of polar issues of South China Business College, Guangdong University of Foreign Studies (Grant No. NGJD 201901) and (Grant No. 20 JD004A).

The significance of pushing up industrial economic cooperation and trade connection with Nordic countries lies in expanding the scale of trade and foreign direct investment, and contributing to promote the development pattern of double circulation and achieving mutually beneficial cooperation for both sides.

Many scholars have focused on the facts that economic globalization makes the industrial chain of a country extending to all over the world and forms a new production model. It has become a common sense in academic circles that the improvement of national industrial economic competitive power depends on the industrial coordination of transnational upstream and downstream, rather than relying on the narrow domestic market and internal limited resources [3]. Particularly, “production segmentation”, the new international production model would affect the position of a country's industries in the global value chain through international trade [4]. Moreover, the integration of industrial chains between the upstream and downstream has been made by the model of “production segmentation” characterized by allocation and transaction of production processes among countries or regions [5].

As the cost of production segmentation involves all costs arising from the separation of specific production links from upstream to downstream, such as costs of tariff, cross-border management and transportation, etc. [6]. Thus, the roles of tariff relief can greatly expand import and export [7] and increase the diversity of trade commodities [8]. The function of trade facilitation is to abate the cost of production segmentation due to cross-border management, and to deepen the division of production between countries [9]. In addition, the improvement of transport infrastructure between trading partners will lead to increasing trade volume [10], enhancing the national position of division of labor in the global value chain indirectly and elevating the international competitiveness of domestic industries [11].

Since the production processes are completed by transnational industrial chain of upstream and downstream, and thus involves the basis of industrial economic cooperation between countries. A number of literatures demonstrate the potentiality of transnational industrial economic cooperation based on trade complementarity which reflects the interdependence of industrial sectors among different countries. The higher the degree of interdependence is, the stronger the basis of industrial cooperation will be.

In the light of the data between China and the 66 countries along “the belt and road”, some scholars [11] found that the strongest trade complementary between both sides is concentrated on resource intensive products. Moreover, the trade complementarity in the above countries is more than trade competition [12]. Bojneč and ferto pointed out that the structural complementarity in different industrial sectors between the two countries would release potential and huge benefits due to the signing of bilateral free trade agreements [13]. In accordance with the calculation of the trade complementarity index between China and Norway, Chen and Lai [14] believe that there is a strong trade complementarity between the two countries. Predictably, it will be a drive force to push up the cooperation both in the upstream and downstream of industrial chains and in the field of high technology as long as establish a free trade area between China and Norway.

The significance of previous literature is associated with pointing out the new production model of industrial connection across boundary, indicating the role of international trade in “production segmentation”, mentioning the possible costs within the processes of “production segmentation” and making efforts towards deal with issues. Yet, how to reduce the cost of “production segmentation” and release the

economic benefits among partner countries has still been a challenge both in theory and practice so far. This article further analyzes the basis and prospects of industrial economic cooperation between China and Nordic countries<sup>①</sup> by means of the calculation of trade complementarity index as per data of international economic organizations and national governments from 2013 to 2018. The purpose of this study is to illustrate that it is possible not only to reduce the segmentation costs in the processes of production through increasing the transactional volume of trade complementary goods and cooperation among trading partners, but also to promote the development of dual circulation pattern in China.

## 2. Methods and Sources

In line with widely used in academia, this paper takes the trade complementarity index as the rationale for demonstrating the basis of industrial cooperation between partner countries. Trade complementarity is defined as the product between the specific exported merchandise with revealed comparative advantage index of country A and the imported goods with revealed comparative disadvantage index of country B [15], which indicates the interdependence of industries and the feasibility of cooperation between both sides. The calculation formula is as follows:

$$TCI_{ij}^k = RCA_{xi}^k \times RCD_{mj}^k$$

Where  $TCI_{ij}^k$  represents the trade complementarity index of merchandise K both in country i and country j,  $RCA_{xi}^k$  means the revealed comparative advantage index of merchandise K in country i,  $RCD_{mj}^k$  shows the revealed comparative disadvantage index of merchandise K in country j and K refers to the specific types of commodity.

Firstly, Revealed Comparative Advantage Index ( $RCA_{xi}^k$ ) is estimated by export (x) in general. The equation is:

$$RCA_{xi}^k = (X_i^k / X_i) / (X_w^k / X_w)$$

Where  $RCA_{xi}^k$  is the revealed comparative advantage index of commodity K in country i,  $X_i^k$  and  $X_w^k$  are the export volume of commodity K from country i and from the world separately,  $X_i$  and  $X_w$  are the total exports of country i and the world as the whole, respectively.

Secondly, the Revealed Comparative Disadvantage Index ( $RCD_{mj}^k$ ) is measured by import (m). The estimated equation is:

$$RCD_{mj}^k = (M_j^k / M_j) / (M_w^k / M_w)$$

Where:

- $RCD_{mj}^k$  — import comparative disadvantage index of commodity K in country j
- $M_j^k$  — import volume of commodity K of country j
- $M_j$  — total import of commodity k of country j
- $M_w^k$  — import volume of commodity K in the world w
- $M_w$  — total import in the world w

Commodity K is determined in accordance with the classification standards for global commodities in the third edition of Standard International Trade Classification (SITC rev.3) of the United Nations. Among the ten categories of commodities classified

by the standard<sup>②</sup>, there are seven of which involve in the trade complementary commodities between China and the four Nordic countries (Norway, Finland, Sweden and Denmark).

Simply, when trade complementarity index value (TCI) is greater than or equal to 1 ( $\geq 1$ ), it means that the commodity K of a country has strong trade complementarity to its partner countries. The larger the index value is, the stronger the trade complementarity will be. When the value of revealed comparative advantage index is greater than 1 ( $RCA > 1$ ), or the value of revealed comparative disadvantage Index is greater than 1 ( $RCD > 1$ ), it indicates that there is a revealed comparative advantage or revealed comparative disadvantage of commodity K in a country.

This research assesses the categorical effects of global trade commodities on the trade complementarity between China and the four Nordic countries in line with data from international and national sources from 2013 to 2018. The data of international trade of commodity classification, total volume of export and import and tariffs in the world, and national trade volume are separately obtained from the database of international trade data of the United Nations (UN Comtrade Database), the world trade organization (WTO) and the bureaus of statistics of the FNCs, the general administration of customs of China and the national bureau of statistics of China.

The time coverage is affected by the UN Comtrade Database availability, of which the key data of classified trade updates every two years in general and the latest data is available as of 2019, and the latest data about China is as of 2018. Nevertheless, this limitation does not imply that the trend and fundamental trade linkage between China and Nordic countries may be changed after 2018 or unable to extent analysis in the field.

### **3. The basis of mutually beneficial cooperation between China and Nordic countries**

This study applies the above formulas to calculate the trade complementarity (TCI), RCA and RCD of specific commodity categories between China and four Nordic countries (FNCs) separately, and demonstrates a stable basis of\_bilateral industrial cooperation.

In accordance with the calculated results, we summarize the characteristics of commodity complementarity between both sides as follows:

First, there is intensive concentration of specific complementary merchandise categories between bilateral trading.

Table 1 shows that Chinese trade complementary goods to the four Nordic countries are clustered on three categories: raw materials, manufactured goods, machinery and transportation equipment, and miscellaneous goods because the average value of trade complementarity index is greater than 1 ( $TCI \geq 1$ ).

**Table 1.** Trade complementarity index of export by China and import from the FNCs (2013-2018)

Types of commodities	EXP. - IMP.	2013	2014	2015	2016	2017	2018	Average value
Manufactured goods classified chiefly by materials (SITC 6)	China-Norway	1.735	1.832	1.686	1.656	1.993	1.718	1.770
	China-Finland	1.293	1.320	1.276	1.273	1.435	1.374	1.329
	China-Sweden	1.462	1.458	1.425	1.407	1.467	1.465	1.447
	China-Denmark	1.598	1.607	1.607	1.619	1.607	1.243	1.547
Machinery and transport equipment (SITC 7)	China-Norway	2.000	1.853	1.684	1.563	1.426	1.426	1.659
	China-Finland	1.357	1.285	1.230	1.277	1.193	1.184	1.254
	China-Sweden	1.777	1.646	1.539	1.469	1.364	1.359	1.526
	China-Denmark	1.533	1.487	1.351	1.298	1.226	1.519	1.402
Miscellaneous goods (SITC 8)	China-Norway	3.296	3.258	2.869	2.853	2.670	2.572	2.920
	China-Finland	2.091	2.039	1.827	1.774	1.732	1.675	1.856
	China-Sweden	1.362	2.093	2.027	1.895	5.541	1.946	2.477
	China-Denmark	3.264	3.368	2.971	2.911	3.024	1.423	2.827

Note: The results of Table 1 is obtained by employing the formula of trade complementarity index to be calculated based on data of international trade data of the United Nations, the world trade organization, the bureaus of statistics of the FNCs, the general administration of customs of China and the national bureau of statistics of China from 2013 to 2018.

Especially, Chinese miscellaneous goods present stronger trade complementarity to Norway, Sweden and Denmark owing to  $TCI > 2.5$  in average value.

As Table 2 reports, the strongest complementary commodities of the FNCs to China are converged on two groups, crude materials (non-edible raw materials excluding fuel) and oils of anim. & veg (SITC 2+4) whose TCI average value of Finland, Sweden and Denmark to China are up to 8.9, 6.6 and 4.5, respectively. The TCI of Norwegian mineral fuels and lubricants (SITC 3) to China are 5.6 in average value. Moreover, the TCI of chemical products (SITC 5) of Denmark (including pharmaceutical products), machinery and transport equipment (SITC 7) of Sweden, and manufactured goods classified chiefly by material of Finland to China are greater than 1.

**Table 2.** Trade complementarity index of export by the FNCs and import from China (2013-2018)

Types of commodities	EXP. - IMP.	2013	2014	2015	2016	2017	2018	Average value
Crude materials + oils of animals & vegetables (SITC 2) + (SITC 4)	Norway-China	1.71	1.94	2.23	2.43	1.50	1.66	1.91
	Finland-China	9.25	9.09	10.42	10.68	7.10	7.09	8.94
	Sweden-China	7.08	7.73	7.55	7.44	5.11	4.77	6.61
	Denmark-China	5.27	5.01	5.47	4.82	3.54	3.09	4.53
Mineral fuels & lubricants (SITC 3)	Norway-China	3.72	4.16	5.56	6.82	6.84	6.42	5.59
Chemical products (SITC 5)	Denmark-China	1.04	1.05	1.36	1.01	1.72	1.69	1.31
Manufactured goods classified chiefly by materials (SITC 6)	Finland-China	1.63	1.81	1.56	1.50	1.40	1.36	1.54
Machinery & transport equip. (SITC 7)	Sweden-China	1.47	1.39	1.39	1.35	1.18	1.17	1.33

Note: The results of Table 2 is obtained by employing the formula of trade complementarity index to be calculated based on data of international trade data of the United Nations, the world trade organization, the bureaus of statistics of the FNCs and the national bureau of statistics of China from 2013 to 2018.

Second, it shows the structurally complementary in bilateral export and import. The industry chain of upstream and downstream, for example, has been formed between both sides. On the one hand, the crude materials (SITC 2) exported by the FNCs to China are the upstream linked with raw material supply chain to Chinese manufactured goods classified chiefly by material (SITC 6). On the other hand, the SITC 6 exported by China is the downstream industrial chain of the FNCs. In addition, the average value of TCI of Chinese miscellaneous goods (SITC 8) to Norway, Sweden, Denmark and Finland are 2.9, 2.5, 2.8 and 1.9 separately (despite fluctuations in different years). It indicates that Chinese export goods (SITC 8), including boxes, bags, shoes, clothing, furniture, heating equipment of housing sanitary water and lighting equipment, can make up for the shortage of the FNCs.

The characteristics are linked with the complementary of intra-industry trade of SITC 7 between China and Sweden, and SITC 6 between China and Finland. The commodities, SITC 7, exported from Sweden to China are mainly the goods of high technology and high added value, while the grade of similar group exported from China to Sweden is relatively low since the average value of the intra-industry trade index (IITI) of SITC 7 in Sweden is 0.97 which is higher than 0.80 in China along with the calculation of the IITI (see Table 3). The average value of the IITI about SITC 6 between China and Finland is close to each other, but Finland is slightly higher, 0.58, than that of China, 0.54 (see Table 4).

**Table 3.** Intra industry trade index of machinery and transportation equipment between China and Sweden (2013-2018)

Countries	2013	2014	2015	2016	2017	2018	Average value
China	0.812	0.807	0.784	0.799	0.809	0.819	0.805
Sweden	0.943	0.973	0.978	0.987	0.983	0.981	0.974

Note: The results of Table 3 is obtained by employing the formula of Intra industry trade index to be calculated based on data of international trade data of the United Nations from 2013 to 2018.

**Table 4.** Intra industry trade index of SITC 6 between China and Finland (2013-2018)

Countries	2013	2014	2015	2016	2017	2018	Average value
China	0.581	0.602	0.511	0.512	0.544	0.547	0.549
Finland	0.553	0.56	0.556	0.582	0.657	0.59	0.584

Note: The results of Table 4 is obtained by employing the formula of Intra industry trade index to be calculated based on data of international trade data of the United Nations from 2013 to 2018.

These results illustrate the transactions of heterogeneous commodities within same category, and shape the structural complementarity in SITC 7 between China and Sweden and SITC 6 between China and Finland, respectively.

Third, the trade complementarity between China and the four Nordic countries is based on their respective comparative advantages. The three major exported categories (SITC 6, SITC 7 and SITC 8) with revealed comparative advantage (RCA) in China are exactly the imported commodities with revealed comparative disadvantage (RCD) in the FNCs as the average value of RCD in those countries are more than 1. The exported merchandises (SITC 2, SITC4 and SITC 3) by the FNCs with revealed comparative advantage (RCA) to China are precisely the revealed comparative disadvantage (RCD) of imported goods in China, in particular, the average value of RCD about SITC 2 + SITC4 is as high as 3.9 which means a strong trade complementary to China (see Table 5).

The characteristics of bilateral trade complementarity are manifested by the connections:(1)The integration of industrial chain between the upstream and the downstream by way of the bilateral transaction of SITC 2 exported by the FNCs and SITC 6 exported by China; (2)structural complementarity of heterogeneity commodities of SITC 7 through intra industry trade between China and Sweden; (3)structural complementarity of heterogeneity goods of SITC 6 by intra industry trade between China and Finland;(4)complementarity between supply and demand via exported chemical merchandises (SITC 5) from Denmark to China.

**Table 5.** Average values of comparative advantage index and comparative disadvantage index between China and the FNCs

Countries	Types of commodities	Average value
Comparative advantage index of China	SITC 6 SITC 7 SITC 8	1.388 1.364 2.115
Comparative disadvantage index of China	SITC 2 + SITC 4 SITC 3 SITC 7	3.989 1.117 1.208
Comparative advantage index of the FNCs	SITC 2 + SITC 4  SITC 3 SITC 7	2.258 (Finland) 1.658 (Sweden) 1.134 (Denmark) 4. 941 (Norway) 1.097 (Sweden)
Comparative disadvantage index of the FNCs	SITC 6  SITC 7  SITC 8	1.275 (Norway) 1.043 (Sweden) 1.547 (Denmark) 1.215 (Norway) 1.118 (Sweden) 1.402 (Denmark) 1.378 (Norway) 1.026 (Sweden) 2.820 (Denmark)

Note: The results of Table 5 is obtained by employing the formulas of revealed comparative advantage index and revealed comparative disadvantage index to be calculated based on data of international trade data of the United Nations, the world trade organization, the bureaus of statistics of the FNCs, the general administration of customs of China and the national bureau of statistics of China from 2013 to 2018.

Generally, the complementary commodities of the FNCs to China cover almost all types of goods, except for SITC 8, in which include not only resource-intensive products (SITC 2+4, SITC 3) and capital intensive goods (SITC 5, SITC 7), but also complementary merchandises between intra-industry (SITC 7) and inter-industry trade (SITC 6). The converge of bilateral complementary commodities is shown that the complementary commodities of China to FNCs are highly concentrated in SITC 6, SITC7 and SITC 8, while the resource intensive goods of the FNCs with strong complementarity to China are mainly centralized in SITC 2 + 4 and SITC 3. The differences of bilateral complementary commodities is reflected by the discrepancy or gap of supply and demand between both sides, in other words, the exporting goods of SITC 6, SITC7 and SITC 8 from China involve in the shortage and demand ones in FNCs and the main exporting goods of SITC 2 + 4 and SITC 3 from FNCs mean the lack and demand of those resources in China. Especially, the complementary groups SITC 2 + 4 of Finland ( $TCI > 8.9$ ), Sweden ( $TCI > 6.6$ ) and SITC 3 ( $TCI > 5.5$ ) of Norway to China are extremely stronger than the rest categories. In addition, the intra-industry trade of SITC 6 between China and Finland presents that the technical level is very close one and other, which indicates the potential competition and implies the urgency of upgrading industrial structure in China.

The characteristics indicate that the industrial economic cooperation between China and the FNCs has a solid basis and sustainability because of the highly interdependent according to their respective export comparative advantages. Therefore, further bilateral cooperation should focus on releasing the trade potential of those commodities, in particular, enhancing the catena between supply chain (SITC 2) and industrial chain SITC 6, and expanding foreign direct investment to Danish chemicals as well.

## 4. Prospect of enhancing bilateral industrial economic cooperation

### 4.1 Expanding bilateral import and export for mutual benefit cooperation

Countries of Norway, Finland, Sweden and Denmark are export-oriented economies which promote free trade policies, and the proportion of export to GDP is close to or more than half [16]. Also, as previously mentioned, the trade complementarity commodities of Nordic countries to China are substantially on resource categories based (SITC 2+ SITC 4, SITC 3). Correspondingly, the demand for industrial resources is increasing rapidly with stimulating economic growth, and thus exacerbates the shortage of domestic resources in China. The imported volume of oil and natural gas, for example, was separately reached 80% and 43% in 2020.

Currently, the efficiency of import from the FNCs is far from meeting Chinese demand for industrial resources. According to the data of international trade published by the United Nations from 2013 to 2017, of the total amount of \$254.5 billion of SITC 2+ SITC 4 imported from the world, the average total amount of imported from the FNCs just accounted for \$2.5 billion that was less than 1% of Chinese total imports and less than one ninth of the total exports of the FNCs. Meanwhile, the total volume of SITC 3 imported from the world was \$250 billion but the import from Norway was less than \$0.5 billion, only one-five-hundredth of Chinese total import and one in 145 of Norwegian total export (\$72.4 billion)<sup>⑨</sup>.

Obviously, this is a positive sum game and mutual benefit for both sides through increasing import and export of complementarity commodities.

### 4.2 Constructing express channel for China railway express

One of the reasons of restricting enlarging bilateral complementarity commodities is linked with small loading capacity and expensive freight by air transport or taking long time by ocean shipping. The price of China railway express (the China - Europe block train ferry) is one fifth lower than that of air transport, and one third less than that of sea shipping. It can avoid the special security risks of traditional ocean routes, such as war or piracy, and make up for the shortcoming of freezing eight months per year of the Arctic channel to be opened in the future.

However, insufficient imported goods in return trip from Nordic countries have led to a high rate of empty box loading of China railway express (CR express). Moreover, different countries have to delay across boundary due to the inconsistency of rail gauge, and thus affect speed and the connection of transport capacity.

In order to deal with this problem, the possible solution is to push on the high-speed railway project of “Moscow-Beijing” under construction between China and Russia<sup>⑩</sup> and to extend it from Moscow to Helsinki based on existing passenger route from Helsinki to Moscow (15 hour one-way arrival). Judging from the feasibility level, Russia lacks capital and technology and is willing to cooperate with China in infrastructure and other investment projects. Finland is inclined to promote the abutting joint between Chinese initiative of “the belt and road” and Finnish “Arctic corridor” plan, so that the transportation network of Finland will become a transportation hub connected with Arctic and Eurasian continent. Chinese production capacity of railway and the technology of high-speed rail rank first in the world. Once open up the high-speed rail line from Beijing to Moscow to Helsinki relying on the cooperation

among China, Russia and Finland, the bottleneck of transport capacity between China and Nordic countries will be significantly alleviated. The commodities of Norway and Sweden will be also increasingly sold to China, Russia and other countries through the freight high-speed rail linked between China and Finland.

### *4.3 Promoting bilateral free trade via institutional arrangements*

#### *4.3.1 Reducing tariff and non-tariff barriers*

Tariff relief between trading partners are conducive to enhancing the international competitiveness of bilateral industries and investment opportunities as well [17]. At present, there are still some tariff and non-tariff barriers between China and the FNCs. Chinese tariffs on trade complementary products remained at a relatively low level, the average tariffs on SITC 3, SITC 2, (SITC 0), and SITC 4 are 5.3%, 6.9%, 10.3%, and 12.4%, respectively, for example. The tariff of Nordic countries was generally low, but there were still various non-tariff barriers and value-added tax were higher than that of China along with the HS classification data of the WTO tariff from 2013 to 2017.

Tariff relief on commodities of trade complementarity can not only increase the mutual benefit in investment and trade, but also contribute to build free trade area network between China and FNCs. On the strength of the institutional arrangement of free trade area, both sides can form a larger market scale and elevate the international competitiveness of member countries [18].

#### *4.3.2 Implementing consistent policies and standards of trade facilitation*

Trade facilitation is the simplification of trade procedures in rules and regulations of market access, e-commerce, customs and port management for the purpose to promote free flow of goods. With decreasing costs of transaction and coordination through trade facilitation, the total trade volume will be expanded while the cost of production segmentation will be reduced between partner countries. For example, the burst of potential trade volume owing to trade facilitation has been more significant than the roles of tariff relief or regional economic integration in the countries of “the belt and road”[19].

At present, the degree of trade facilitation in China and the FNCs needs to be improved. Based on the survey, the lingering situation of CR express caused by harmonizing various transport routes and conducting customs inspection will be dropped by at least 40% through promoting unified conventions of railway transport and integrating standards of customs inspection among countries of CR express passed through [20].

## **5. Conclusion**

With increasing industrial interdependence among countries, the depth and breadth of cooperation among partner countries are bound to be promoted and expanded under economic and production globalization. The trade complementarity between China and FNCs provides a stable basis for bilateral cooperation and development space. Moreover, the cooperation based on win-win result shows the common interests in the

long term because bilateral industrial economies depend on each other and economic interests are integrated. However, trade barriers caused by tariffs, inconsistent regulations of customs formalities and imperfect transport infrastructure have pushed up transaction costs and limited the trade expansion of bilateral complementary commodity.

For the sake of setting off domestic circulation, meeting the demand of resources and decreasing costs of trade and production segmentation, it is urgent to promote external circulation with Arctic countries. The further cooperation should focus on reducing tariffs, improving transportation infrastructure, pursuing the common policies and standards for trade facilitation, and promoting free trade zone so that release potentiality of trade and investment.

The aim of enhancing the industrial economic cooperation between China and the FNCs is to promote more open world economy and to create new advantages in international cooperation and competition.

**Acknowledgments:** The authors acknowledge financial support from the Education Bureau of Guangdong Province of China, the research center of polar issues of South China Business College, Guangdong University of Foreign Studies, and reviews and editors' valuable instructions for revision of this manuscript. The authors are responsible for all errors.

## Notes

- ① The Nordic countries mentioned in this paper only include Norway, Finland, Sweden and Denmark because Iceland established a free trade zone with China in 2013. Thus, this study does not include Iceland.
- ② As some commodity categories are merged in international trade database about the commodity classification of the United Nations: Food and live animals (SITC 0) + beverages and cigarettes (SITC 1), crude raw materials (SITC 2) + animal and vegetable oils and waxes (SITC 4), this paper also takes the merged data. In addition, unclassified commodities (coins: non legal tender, non-monetary gold excluding gold ore, SITC 9) are not widely traded commodities and thus are not covered into the analysis of this paper.
- ③ According to the international trade data of the United Nations from 2013 to 2017, authors calculated the average volume of import and export of SITC 2, SITC 4 and SITC 3.
- ④ In accordance with Russian media report, the project of "Moscow-Beijing" high-speed railway aimed at building the Eurasian high-speed transportation corridor was determined by signing a memorandum of understanding on high-speed railway transportation between Chinese railway construction corporation and the Ministry of transport of the Russian Federation, and between Chinese Commission of development and reform and Russian Railway Corporation in October 2014. Chinese side would provide technology for the implementation of the project and for the planning and construction  
(Source: The Chinese will build a subway station in Moscow, *True, Russia*, January 25, 2017).

## References

- [1] Liu H. Speeding up establishing a new development pattern of double circulation with domestic circulation as the main stream and mutual promotion between domestic and international circulation. People's Daily, Nov; 25, 2020.
- [2] Chinese central government. China's Arctic Policy. The State Council Information Office of the People's Republic of China, Jan; 2018.
- [3] Kong L C, Li Y. On the new trend of industrial economic development, Research on Financial Issues. 2003 (09): 56-63.
- [4] Liu QL, Gao Y, Han JW. Productivity effect of international production segmentation, Journal of Economic Research. 2010 (02): 27-41.

- [5] Sun PY, Zhang CY, Yang YQ. Production segmentation, information friction and tariff transmission: Theory and experience of consumer market. *World Economy*. 2021 (02): 16-32.
- [6] Wang YN, Zhou MT. Trade policy uncertainty, tariff reduction and distribution of export products. *Quantitative Economic and Technological Economic Research*. 2017 (12): 29-43.
- [7] Aditya A, Acharyya R. Trade Liberalization and Export Diversification. *International Review of Economics and Finance*. 2015(39):390-410.
- [8] Hummels DL. Transportation costs and international trade in the second era of globalization. *Journal of Economic Perspectives*. 2007, 21(3):139.
- [9] Chen H, Liu JY. one belt, one road, the impact of infrastructural development on China's foreign trade: Based on panel threshold model. *International Business*. 2020 (04): 51-65.
- [10] Wang YY, Tu MH. Can production segmentation improve the position of global value chain division? An empirical analysis based on domestic and international double cycle theory. *China's Industrial Economy*. 2021 (02): 29-46.
- [11] Hu Y, Yan JL, Quan Y. Measurement of trade complementarity between China and countries along the "21st century Maritime Silk Road" and its influencing factors. *World Economic Research*. 2017 (08): 20-39.
- [12] Li J, Chen N, Wan GH, Chen S. The complementary relationship and dynamic change of goods trade along the "one belt and road" -- Based on the network analysis method. *Management World*. 2017 (04): 37-55.
- [13] Bojnec S, Ferto I. Complementarities of trade advantage and trade competitiveness measures. *Applied Economics*. 2012 (44):399–408.
- [14] Chen YH, Lai MY. Analysis of bilateral trade structure and competition complementarity between China and Norway, *Financial Theory and Practice*, 2011 (09): 22-33.
- [15] Drysdale P. Japan, Australia and New Zealand: the Prospects for Western Integration. *Economic Record*. 1969 (45):321-342.
- [16] Zheng YQ. Jointly building blue economic corridor between China and Northern Europe: foundation and challenge. *Research on International Issues*. 2019 (12): 14-28.
- [17] Li G, Ye X. Discussion on the rationality of China's tariff level and tariff structure under the new situation. *International Trade Issues*. 2017 (07): 39-54.
- [18] Stack M.M., Bliss M. EU economic integration agreements, exit and trade. *Review of World Economics*. 2020 (5):1-31.
- [19] Kong QF, Dong HY. Trade facilitation level measurement and trade potential research of countries along "the belt and road". *International Trade Issues*. 2015 (12): 62 -78.
- [20] Kang YF. Research on improving the operation speed of China -Europe block train. *Logistics Engineering and Management*. 2016 (8): 22-31.

# Tariff Relief and Promoting Free Trade Agreement Between China and Norway

Nianlin Zhou<sup>a1</sup> Jin Guo<sup>b</sup> and Manyuan Jiang<sup>a</sup>

<sup>a</sup>South China Business College, Guangdong University of Foreign Studies,  
Guangzhou, China

<sup>b</sup>College of Management, Guangdong AIB Polytechnic, Guangzhou, China

**Abstract.** Tariff relief is a prerequisite for reaching a free trade agreement. This study employs quantitative indicators to identify the specific types of goods with trade complementarity between China and Norway based on the published data by the United Nations, the World Trade Organization and the National Bureaus of statistics both in China and Norway. The empirical results confirm that some types of commodities are provided with trade complementarity between both sides. Nevertheless, these complementary goods are imposed tariffs on each other. The consequences are linked with aggravating resources shortage in China on the one hand, and limiting consumption in Norway on the other. Therefore, pushing up tariff relief is favor of mutual benefit cooperation and making progress in the negotiation of free trade agreement between China and Norway.

**Key words.** Tariff Relief; free trade agreement; trade complementarity

## 1. Introduction

Free trade agreement is the key to promote free trade zone. The establishment of free trade zone is usually based on the signing of free trade agreement (FTA). Recent years, FTA has become an important paradigm to enhance economic and trade cooperation between China and countries along “the belt and road”.

As the part of “the belt and road”, the Nordic countries have been pursuing free trade policy for a long time and have expressed strong complementarity with Chinese industries. Iceland has established a free trade zone with China in 2013. The negotiations of free trade agreement between China and Norway have conducted the 16th round so far.

The fundamental content of free trade agreement is tariff relief [1]. The important premise of reaching FTAs involves the types, range and size of goods for tariff relief. Moreover, the expansion of bilateral trade volume caused by tariff relief means extending free trade further and increasing the possibility to reach a FTA for both sides.

Many studies have provided a lot of valuable research on the role and effect of free trade zone. The free trade zone is defined as the elimination of tariff and non-tariff

---

<sup>1</sup> Corresponding Author: Nianlin Zhou, South China Business College, Guangdong University of Foreign Studies, Guangzhou, China; Email: gweduyj@163.com. This research was funded by the Education Bureau of Guangdong Province of China (Grant No. 2018GXJK272) and the research center of polar issues of South China Business College, Guangdong University of Foreign Studies (Grant No. NGJD 201901) and (Grant No. 20 JD004A).

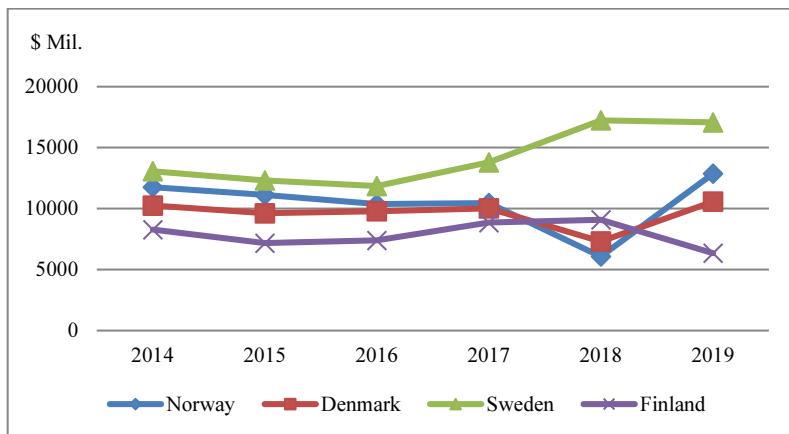
barriers between countries or regions through the signing of free trade agreements [2]. Countries with high production costs will benefit from the expansion of trade size and trade creation effect [3]. Moreover, free trade zone (FTZ) is a new way of economic and political interdependence that promotes the combination of national or regional development strategy and international process. The most direct goal of the FTZ is to expand the mutual trade among member countries and to form a larger market scale so that enhance international competitiveness [4]. In addition, some scholars have paid attention to the effects of establishing FTZ between China and other countries. China and ASEAN Free Trade Area, for example, the trade flow between member countries with similar per capita income and demand structure might increase [3]. The income distribution of the Asia Pacific free trade area is different among member countries. The developed countries, represented by the United States, get more additional income than the developing countries [5].

Others find that trade complementarity is an important way to promote the FTZ [6]. The difference of resource endowment has become the decisive factor of trade complementarity between China and the countries or regions along “the belt and road” [7]. The complementarity of agricultural product transaction between China, Norway and Finland is weak, but is strong with Sweden, Denmark and Iceland [8]. If establish a FTZ, it will enhance the cooperation in the field of high technology due to the complementarity of inter industry trade between China and Norway [9]. Generally, the potential and huge benefits will be stimulated by signing the free trade agreement (FTA) as long as there is industrial structural complementarity between trade partners [10]. However, the average import efficiency from the countries along “the bet and road” is lag behind the average level of export from China [11].

In terms of tariff relief, the basic content of reaching a FTA, almost all of the FTA in the world have covered the tariff relief for manufactured and agricultural products [12]. Since tariff reduction will make contribution to significantly increase the volumes of export and import products [1] [13], and to add the diversity of trade goods as well [14].

Current studies discuss the macro-economic effect in building the FTZ, but lack deep analysis about the influence of tariff relief on import volume and consumption demand with complementary goods in bilateral trade and on the FTA.

Particularly, China has become the eighth largest export market and the fifth largest import source of Sweden, the seventh largest export market and the fourth largest import source of Denmark, the third largest source of imports and ninth largest export market of Norway, and the fifth largest import and export market of Finland in 2019 according to the data of Eurostat. Moreover, compared with the major Nordic countries, Norway has become an important trading partner with China, second only to Sweden (see Figure 1).



**Figure 1.** The total volume of import and export of Norway, Denmark, Sweden and Finland to China (2014-2019)

**Source:** The results of Figure 1 is obtained as per the international trade data of the United Nations, Eurostat and general administration of customs of China.

How to further promoting economic cooperation between Norway and other Nordic countries should be the focus of research. The possible contribution of this article is to identify the commodities with trade complementarity through applying the trade complementarity index in accordance with data of international economic organizations and national statistics, and to analyze the impact of tariff on the import volume of complementary products for both sides. The suggestion is to decrease or to abolish tariffs on goods with strong complementarity first, so that take it as a breakthrough to promote new progress in FTA negotiation between China and Norway.

## 2. Identifying the commodities of trade complementarity between China and Norway

It is the basic condition for promoting the mutual benefit and positive sum game by means of identifying the types of complementary goods between China and Norway. This study classifies all commodities according to the classification standard of the United Nations International Trade Classification (SITC Rev. 3) (see Table 1), so as to clarify the specific commodity types related to the trade complementarity between both sides, and avoid too general description about complementary goods to carry out policy measures.

**Table 1.** The standard of international trade classification of United Nations (SITC Rev. 3)

Classification of SITC Rev.3	Description of classified products
SITC 0	Food and live livestock
SITC 1	Beverages and cigarettes
SITC 2	Crude raw materials (non-edible raw materials other than fuel, including: raw fur, rubber, oilseeds, wood cork, pulp waste paper, silk, cotton hemp textile fiber, sand and gravel abrasive, pyrite, iron copper nickel aluminum ore and concentrate)
SITC 3	Fossil fuels and raw materials
SITC 4	Fats and waxes of animal and vegetable
SITC 5	Chemicals
SITC 6	Manufactured goods classified chiefly by materials (leather products, textiles, non-ferrous metals, steel products, building materials, glass ceramic products, paper and paperboard products)
SITC 7	Machinery and transport equipment
SITC 8	Miscellaneous goods (luggage, shoes, clothing and furniture, lighting and photographic equipment, house sanitation and water heating devices, etc.)
SITC 9	Unclassified commodities and transactions (coins-non legal tender, non-monetary gold-excluding gold ore)

**Source:** The international trade database of United Nations.

**Note:** As SITC 0 + 1 and SITC 2 + 4 are consolidated in the statistical classification of the above database, and thus SITC 0 + 1 and SITC 2 + 4 are combined in the following calculation in this paper. In addition, SITC 9 is not included in the study because it is not a widely traded commodity.

## 2.1 The 'revealed' comparative advantage of China and Norway

The index of 'revealed' comparative advantage (RCA), an index for evaluating export comparative advantage and its international competitiveness of a country, is selected to measure the export comparative advantage for both countries. The formula of revealed comparative advantage index is:

$$RCA^A_{xi} = (X^A_i/X_i)/(X^A_w/X_w) \quad (1)$$

Where  $RCA^A_{xi}$  means the export comparative advantage index of country  $i$  in commodity  $A$ , and  $X^A_i$  and  $X^A_w$  separately involve the export volume of country  $i$  and the world in product  $A$ ,  $X_i$  and  $X_w$  express the total export of country  $i$  and the world, respectively. When the revealed comparative advantage index (RCA) is more than or equal to 2.5, it illustrates that the export competitiveness of this commodity is very strong. If the  $RCA > 1$ , the commodity has a comparative advantage in world exports, otherwise, it is weak.

The calculation results of RCA formula is shown by Table 2 in which presents that Chinese products with export comparative advantage for Norway include three categories: raw material products (SITC 6), machinery and transportation equipment (SITC 7) and miscellaneous products (SITC 8). Because the critical value of RCA is greater than or equal to 1.

As per the trend of export comparative advantage from 2008 to 2017, the comparative advantage of SITC 8 increased rapidly since the index value of RCA of SITC 8 was more than 2. The RCA of SITC 7 fluctuated slightly but its average value still remained at 1.44. The comparative advantage of SITC 6 was relatively stable due to the average value of RCA, 1.35. The RCA value of other commodities was less than 1, which meant no export comparative advantage.

**Table 2.** The revealed comparative advantage of export goods in China (2008—2017)

Year	SITC 0+1	SITC 2+4	SITC 3	SITC 5	SITC 6	SITC 7	SITC 8
2008	0.224722	0.200551	0.015612	0.461114	0.798843	1.65053	1.653093
2009	0.419008	0.191114	0.123615	0.461997	1.256529	1.477847	1.451291
2010	0.43905	0.178183	0.113664	0.516612	1.258859	1.484331	1.61573
2011	0.444462	0.181205	0.099113	0.578156	1.333935	1.507345	1.857
2012	0.424401	0.16992	0.085027	0.538652	1.369643	1.492561	2.258569
2013	0.402386	0.162679	0.089499	0.5237	1.376928	1.473915	2.313704
2014	0.390203	0.174482	0.0928	0.547597	1.416243	1.390803	2.27559
2015	0.38557	0.169204	0.109167	0.522596	1.403972	1.315355	2.059561
2016	0.41776	0.169795	0.138632	0.519241	1.381799	1.272734	2.015575
2017	0.414315	0.164022	0.150758	0.574912	1.359937	1.338612	2.022959

**Source:** The results of Table 2 is obtained by using the formula of RCA to calculate according to data of international trade data of the United Nations, the world trade organization and the national bureau of statistics of China from 2008 to 2017.

Over the same period, there were two types of resource intensive products expressed remarkable export comparative advantage in Norway (see Table 3). SITC 3, for example, was very strong with more than 4 of the RCA value and reached 5.79 in 2016. The export comparative advantage of SITC 0 + 1 was gradually increased and the average value of RCA was 1.2. However, the RCAs of the rest did not indicate export competitive advantage ( $RCA < 1$ ).

**Table 3.** The revealed comparative advantage of export goods in Norway (2008—2017)

Year	SITC 0+1	SITC 2+4	SITC 3	SITC 5	SITC 6	SITC 7	SITC 8
2008	0.755146	0.283107	3.960044	0.245600	0.653422	0.279400	0.159072
2009	0.906693	0.302748	4.724704	0.286198	0.635979	0.338554	0.156100
2010	1.031400	0.381097	4.655100	0.31011	0.70976	0.282005	0.147842
2011	0.993666	0.324286	3.952567	0.275335	0.623594	0.248177	0.176155
2012	0.904495	0.369216	3.900466	0.218909	0.574747	0.250705	0.213723
2013	1.094083	0.382814	3.932067	0.229946	0.584351	0.273153	0.226455
2014	1.190636	0.437176	4.065533	0.256974	0.609723	0.290087	0.235483
2015	1.310207	0.500147	5.174210	0.302896	0.698820	0.328463	0.261614
2016	1.754480	0.526113	5.794143	0.350653	0.735263	0.324753	0.268368
2017	1.637383	0.474539	5.513847	0.332156	0.734788	0.270672	0.236565

**Source:** The results of Table 3 is obtained by using the formula of RCA to calculate according to data of international trade data of the United Nations, the world trade organization and the national bureau of statistics of Norway from 2008 to 2017.

## 2.2 The ‘Revealed’ import comparative disadvantage of China and Norway

The index of ‘revealed’ comparative disadvantage (RCD) is widely applied to measure the comparative disadvantage in producing or importing specific commodity of a country. The revealed comparative disadvantage index (RCD) can be expressed as:

$$RCD^A_{mj} = (M^A_j/M_j)/(M^A_w/M_w) \quad (2)$$

Where  $RCD^A_{mj}$  represents the import comparative disadvantage index of country j for goods A,  $M^A_j$  is the import volume of A commodity of country j,  $M_j$  means the total import volume of country j,  $M^A_w$  refers to the import volume of world of product A, and  $M_w$  is the total import volume of the world. If  $RCD^A_{mj} > 1$ , it indicates a comparative disadvantage both in the production and import of product A in country j.

Table 4 and table 5 are the results of using the formula of RCD to measure the import comparative disadvantage of China and Norway, respectively. There are three types of goods with import comparative disadvantage in China, including SITC 3, SITC 2 + 4 and SITC 7 (see Table 4). Among them, SITC 2 + 4 have the biggest import comparative disadvantage owing to its RCD is more than 4, even though declines after 2016, still higher than 3.16. During the nine-year period from 2009 to 2017, Chinese import demand for SITC 3 and SITC 7 was strong ( $RCD > 1$ ) which implied the

shortage of those commodities in China.

**Table 4.** The revealed comparative disadvantage of import goods in China (2008—2017)

Year	SITC 0+1	SITC 2+4	SITC 3	SITC 5	SITC 6	SITC 7	SITC 8
2008	0.590681	0.29527	0.209935	0.528153	0.93832	1.805158	1.364172
2009	0.244802	5.010623	0.887799	0.972409	0.891796	1.373388	0.504159
2010	0.277571	4.490528	0.900858	0.96212	0.76731	1.338907	0.562489
2011	0.294669	4.328462	0.918676	0.941401	0.699318	1.278152	0.576681
2012	0.35039	4.343187	0.969233	0.920865	0.698295	1.263373	0.695401
2013	0.36741	4.455592	0.947198	0.900874	0.663855	1.270136	0.673148
2014	0.393886	4.426294	1.022617	0.894669	0.747819	1.257872	0.663857
2015	0.480798	4.466419	1.074015	0.890647	0.676215	1.294549	0.690631
2016	0.487454	4.621081	1.177042	0.883231	0.656253	1.274892	0.673933
2017	0.472035	3.161206	1.24079	0.917028	0.636521	1.093867	0.645583

**Source:** The results of Table 4 is obtained by using the formula of RCD to calculate as per data of international trade data of the United Nations, the world trade organization and the national bureau of statistics of China from 2008 to 2017.

Table 5 reports that five categories of products of Norway have import comparative disadvantage. Based on the order from high to low of import comparative disadvantage in 2017, SITC 6 has the highest import comparative disadvantage, followed by SITC 2 + 4, SITC 8, SITC 7 and SITC 0+1. The index values of RCD for those group goods are all more than 1, and thus confirm there are a strong import demand for these products, especially for SITC 6 in Norway.

**Table 5.** The revealed comparative disadvantage of import goods in Norway (2008—2017)

Year	SITC 0+1	SITC 2+4	SITC 3	SITC 5	SITC 6	SITC 7	SITC 8
2008	1.069645	2.529438	0.281721	0.852741	1.252265	1.36969	1.009783
2009	1.082072	2.20536	0.358126	0.865792	1.25719	1.350053	0.932769
2010	1.116182	2.155827	0.436114	0.865119	1.198074	1.307791	1.026885
2011	1.125976	2.112751	0.370559	0.834193	1.198267	1.355258	1.139269
2012	1.221223	2.026445	0.342149	0.843441	1.320951	1.342006	1.377953
2013	1.226927	1.897206	0.41185	0.826468	1.259975	1.356997	1.424652
2014	1.223017	2.026531	0.334176	0.834644	1.29336	1.332163	1.431687
2015	1.175403	2.211603	0.393308	0.801059	1.200599	1.280051	1.393118
2016	1.20437	2.182406	0.377018	0.839804	1.198277	1.228296	1.415325
2017	1.099662	1.391913	0.410331	0.843258	1.465317	1.065188	1.319775

**Source:** The results of Table 5 is obtained by using the formula of RCD to calculate based on data of international trade data of the United Nations, the world trade organization and the national bureau of statistics of Norway from 2008 to 2017.

### 2.3 Trade complementarity between China and Norway

Trade complementarity index (TCI) is a quantitative index to measure the degree of trade complementarity among trading partner countries. The calculation formula is as follows:

$$TCI_{ij}^A = RCA_{xi}^A \times RCD_{Mj}^A = (X_i^A / X_j) / (X_w^A / X_w) \times (M_j^A / M_i) / (M_w^A / M_w) \quad (3)$$

Where  $TCI_{ij}^A$  means trade complementarity index of country i and country j in commodity A,  $RCA_{xi}^A$  indicates trade comparative advantage of country i in exporting product A,  $RCD_{Mj}^A$  refers to comparative disadvantage of country j in importing merchandise A. When  $TCI > 1$ , it expresses the existence of the trade complementarity in product A between country i and country j, instead, it is low.

Table 6 presents that SITC 6, SITC 7 and SITC 8 of China have strong trade complementarity to Norway ( $TCI > 1$ ). Especially, the group of SITC 8 shows a stronger trade complementarity to Norway, while the trade complementarity SITC 6 is on the rise.

**Table 6.** Trade complementarity goods of export by China and import from Norway (2008-2017)

Year	SITC 0	SITC 1	SITC 2+4	SITC 3	SITC 5	SITC 6	SITC 7	SITC 8
2008	0.016588	0.011737	0.083593	0.831352	0.129714	0.613119	0.504361	0.217002
2009	0.024524	0.013012	1.516956	4.194587	0.278302	0.567164	0.464966	0.078699
2010	0.034392	0.010985	1.711327	4.193584	0.298363	0.544606	0.377578	0.083159
2011	0.030971	0.012704	1.403660	3.631128	0.259201	0.436091	0.317208	0.101585
2012	0.038938	0.017231	1.603574	3.78046	0.201586	0.401343	0.316734	0.148623
2013	0.043629	0.015579	1.705663	3.724446	0.207152	0.387924	0.346941	0.152438
2014	0.053370	0.020177	1.935070	4.157483	0.229907	0.455962	0.364892	0.156327
2015	0.073315	0.000004	2.233866	5.557179	0.269773	0.472553	0.425211	0.180679
2016	0.096144	0.000006	2.431211	6.819950	0.309708	0.482519	0.414025	0.180862
2017	0.094740	0.000004	1.500116	6.841526	0.304596	0.467708	0.296079	0.152722

**Source:** The results of Table 7 is obtained by employing the formula of TCI to calculate based on data of international trade data of the United Nations, the world trade organization, the bureau of statistics of Norway, the national bureau of statistics of China and the national bureau of statistics of Norway from 2013 to 2018.

The Norwegian goods that can meet Chinese consumption demand involve SITC 3 and SITC 2+4 (see Table 7). The TCI value of SITC 3 had been gradually rising from 4.9 of 2009 to 6.8 of 2017. In addition, Chinese import demand about SITC 0 is increasing. The prospects of complementary supply and demand of food products will be very broad as Norway is the largest aquaculture country for aquatic products, and China is one of the largest consumer countries of aquatic products in the world.

**Table 7.** Trade complementarity goods of export by Norway and import from China (2008-2017)

Year	SITC 0	SITC 1	SITC 2+4	SITC 3	SITC 5	SITC 6	SITC 7	SITC 8 <sup>(d)</sup>
2008 <sup>(c)</sup>	0.038637 <sup>(c)</sup>	0.007875 <sup>(c)</sup>	0.507281 <sup>(c)</sup>	0.004398 <sup>(c)</sup>	0.393211 <sup>(c)</sup>	1.000363 <sup>(c)</sup>	2.260714 <sup>(c)</sup>	1.669265 <sup>(c)</sup>
2009 <sup>(c)</sup>	0.045123 <sup>(c)</sup>	0.013362 <sup>(c)</sup>	0.421475 <sup>(c)</sup>	0.04427 <sup>(c)</sup>	0.399993 <sup>(c)</sup>	1.579696 <sup>(c)</sup>	1.995172 <sup>(c)</sup>	1.353719 <sup>(c)</sup>
2010 <sup>(c)</sup>	0.049361 <sup>(c)</sup>	0.008929 <sup>(c)</sup>	0.384132 <sup>(c)</sup>	0.04957 <sup>(c)</sup>	0.446931 <sup>(c)</sup>	1.508206 <sup>(c)</sup>	1.941195 <sup>(c)</sup>	1.659169 <sup>(c)</sup>
2011 <sup>(c)</sup>	0.050176 <sup>(c)</sup>	0.014801 <sup>(c)</sup>	0.382841 <sup>(c)</sup>	0.036727 <sup>(c)</sup>	0.482294 <sup>(c)</sup>	1.59841 <sup>(c)</sup>	2.042841 <sup>(c)</sup>	2.115623 <sup>(c)</sup>
2012 <sup>(c)</sup>	0.057412 <sup>(c)</sup>	0.017234 <sup>(c)</sup>	0.344334 <sup>(c)</sup>	0.029092 <sup>(c)</sup>	0.454321 <sup>(c)</sup>	1.809231 <sup>(c)</sup>	2.003026 <sup>(c)</sup>	3.112202 <sup>(c)</sup>
2013 <sup>(c)</sup>	0.048574 <sup>(c)</sup>	0.016046 <sup>(c)</sup>	0.308636 <sup>(c)</sup>	0.03686 <sup>(c)</sup>	0.432821 <sup>(c)</sup>	1.734895 <sup>(c)</sup>	2.000098 <sup>(c)</sup>	3.296223 <sup>(c)</sup>
2014 <sup>(c)</sup>	0.048791 <sup>(c)</sup>	0.018562 <sup>(c)</sup>	0.353593 <sup>(c)</sup>	0.031012 <sup>(c)</sup>	0.457049 <sup>(c)</sup>	1.831712 <sup>(c)</sup>	1.852776 <sup>(c)</sup>	3.257933 <sup>(c)</sup>
2015 <sup>(c)</sup>	0.046932 <sup>(c)</sup>	0.019351 <sup>(c)</sup>	0.374212 <sup>(c)</sup>	0.042936 <sup>(c)</sup>	0.41863 <sup>(c)</sup>	1.685607 <sup>(c)</sup>	1.683721 <sup>(c)</sup>	2.869212 <sup>(c)</sup>
2016 <sup>(c)</sup>	0.050808 <sup>(c)</sup>	0.022356 <sup>(c)</sup>	0.370562 <sup>(c)</sup>	0.052267 <sup>(c)</sup>	0.436061 <sup>(c)</sup>	1.655778 <sup>(c)</sup>	1.563294 <sup>(c)</sup>	2.852694 <sup>(c)</sup>
2017 <sup>(c)</sup>	0.40074 <sup>(c)</sup>	0.163485 <sup>(c)</sup>	0.228304 <sup>(c)</sup>	0.061861 <sup>(c)</sup>	0.484799 <sup>(c)</sup>	1.992739 <sup>(c)</sup>	1.425873 <sup>(c)</sup>	2.669851 <sup>(c)</sup>

**Source:** The results of Table 6 is obtained by employing the formula of TCI to calculate based on data of international trade data of the United Nations, the world trade organization, the bureau of statistics of Norway, the general administration of customs of China and the national bureau of statistics of China from 2013 to 2018.<sup>(c)</sup>

The complementary product structure illustrates that, on the one hand, Norwegian commodity with strongest export competition is SITC 3. The export comparative advantage of SITC 0 + 1 needs to be supported by importing such goods. Furthermore, there are five categories with import comparative disadvantage, excepting SITC 3 and SITC 5. Thus, the export comparative advantage relies on a single category, SITC 3, while scarce goods highly depend on foreign import in Norway. On the other hand, Chinese export comparative advantages are concentrated on SITC 6, SITC 7 and SITC 8, and the import comparative disadvantages focus on SITC 2 + 4 and SITC 3 which indicate a large resource gap in China.

### 3. Characteristics and tariffs of trade complementary commodities in China and Norway

#### 3.1 Characteristics of trade complementary goods in China and Norway

The product gradation of trade complementary provides that resource intensive products of Norway are major complementarity categories with China, such as fossil

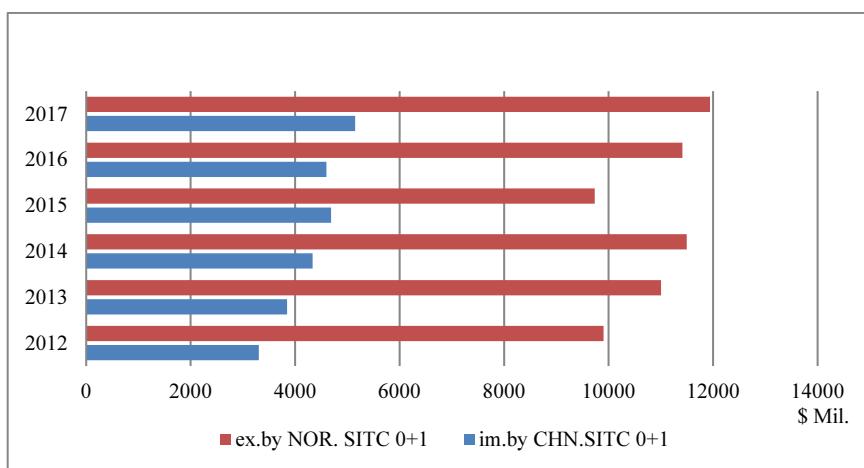
fuels. Meanwhile, some goods of China, including capital intensive products of SITC 6 and SITC 7 and labor-intensive goods of SITC 8, are mainly complementarity products with Norway. The transaction between both sides is linked with inter-industry trade rather than intra-industry complementary trade.

In the light of the types of potential trade complementarity, SITC 2 of Norway and SITC 6 of China have prominent complementarity between the upper and lower industrial chains, which shows the characteristics of industrial division of labor with strong complementarity, and belongs to the type of potential complementarity enhancement. The bilateral trade volume of SITC 2 and SITC 6 should be greatly increased in the long run.

SITC 7 of China touches upon the type of potential expansion because problems of infrastructure construction in Norway where 615 000 buildings are in the high decay hazard category, including Oslo and Hordaland. Moreover, the backlog in road, railway and offshore infrastructure maintenance make the operation of infrastructure extremely severe[15]. The situation of Norwegian infrastructure has not been significantly improved. It is possible for China to increase export of SITC 7 to Norway when familiar with the technical quality standards of Norwegian infrastructure and improve the quality accordingly.

The transaction of SITC 3 between Norway and China has great potentiality. The TCI of SITC 3 with China was as high as 6.8 in 2017. It means that China has strong import potential and huge market demand for this commodity.

The goods of SITC 0 refer to the type of potential excavation. Norway is the second largest exporter of live livestock and fish among the four Nordic countries (Denmark, Finland and Swede), of which salmon and sea cucumber are the scarce species for China. Yet, the average export of SITC 0+1 of Norway was total US \$10.9 billion, but Chinese import from Norway was less than US \$4.3 billion in average from 2012 to 2017 (see Figure 2). Comparing with imported \$15.7 billion from the world in the same period, the import of SITC 0+1 from Norway was around one fourth.



**Figure 2.** The transaction of SITC 0+1 exported by Norway and imported by China (2012-2017)

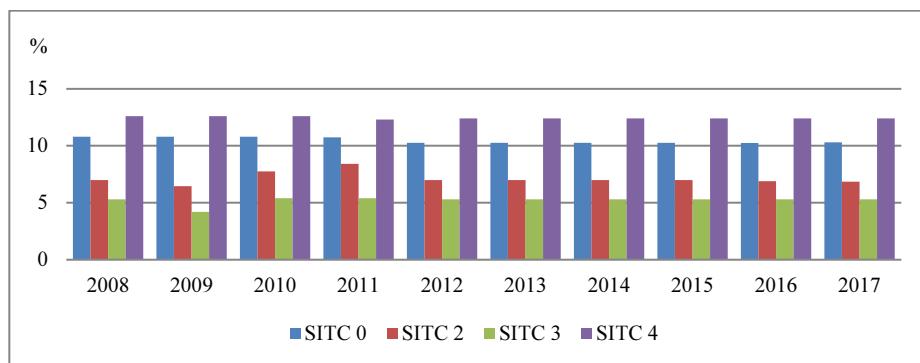
**Source:** The results of Figure 2 is obtained by calculating based on the international trade data of the United Nations, the world trade organization, the bureau of statistics of China and the bureau of statistics of Norway.

The characteristics of bilateral complementarity demonstrate a low competitiveness, high complementarity, the export advantage of one party being the import disadvantage of the other, strong stability and fast growth. In 2020, the total trade volume between China and Norway was US \$16.175 billion, an increase of 24.04% over 2019. China has become the fourth largest trading partner accounting for 9.9% of Norwegian total goods trade [16].

According to the characteristics of bilateral complementarity, China should expand international trade cooperation with Norway in the following fields: energy and fishing resources, complementary sectors in upstream and downstream industrial chains, and infrastructure linked with machinery and transportation equipment manufacturing. However, the expansion of trade potentiality will be negatively affected by the overall bilateral tariff level. Instead, reaching free trade agreements and improving infrastructure will greatly improve the efficiency of bilateral trade.

### *3.2 Tariffs imposed by China on Norwegian trade complementary products*

Chinese tariffs on Norwegian trade complementary products remained at a relatively low level during the period from 2001 to 2017. The average tariffs on SITC 3, SITC 2, SITC 0, and SITC 4 are separately 5.3%, 7.1%, 10.3%, and 12.4% (see Figure 3).



**Figure 3.** The tariff rate imposed by China on Norwegian trade complementary products (2008-2017)  
Source: The results of Figure 3 is obtained by classifying and calculating based on the HS tariff classification data of the WTO.

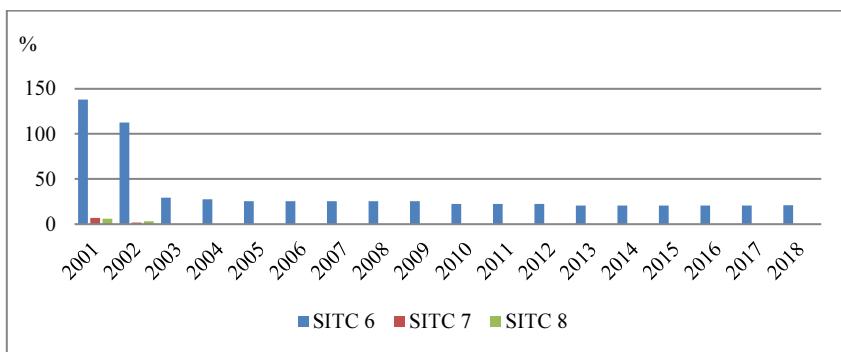
The TCI of Norwegian SITC 2 + 4 to China is 1.6 on average, which implies a strong demand for such products in China. The key constraint to Chinese economic development in the coming period is the shortage of resources. Imposing tariffs on commodities with strong resource complementarity, such as SITC 2 and SITC 0 of Norway, will exacerbate resource shortage in China.

### *3.3 Tariffs imposed by Norway on Chinese trade complementary products*

The imposed tariffs on Chinese SITC 6 has been kept at 21%, even though the tax rate on SITC 6 decreased from 137.9% in 2001 to 21% in 2017 in Norway. The Norwegian tariffs on Chinese SITC 7 and SITC 8 has implemented zero tariffs since 2003 (see Figure 4). Nevertheless, non-tariff barriers, technical standards, and value-added tax

were higher than that of China<sup>2</sup>.

The TCI of Chinese SITC 6 with Norway is as high as 2, and forms an interdependent industrial chain between upstream and downstream in the production process. The results of imposing tariffs on SITC 6 would definitely impact on Norwegian import and consumption demand. Tariff reduction between both sides is the key content of reaching a free trade agreement. Chinese government has been actively promoting signing the free trade agreements. Norwegian foreign minister and finance minister said that “reaching a free trade agreement with China is the top priority of the Norwegian government” [17], and “the tariff agreement is an important prerequisite for trade and investment between Norway and China. Therefore, both sides should consult on it as soon as possible”[18]. In fact, it has always pursued a free trade policy and has independently reduced tariffs for many times in China.



**Figure 4.** The tariff rate imposed by Norway on Chinese trade complementary goods (2008-2017)

**Source:** The results of Figure 4 is obtained by classifying and calculating based on the HS tariff classification data of the WTO.

Currently, the world trade pattern, including China, has been changing, that is, promoting free trade zones or regional super large free trade agreements as an important paradigm to enhance economic cooperation among countries, which requires reducing tariffs and expanding exports and imports. The important part of foreign trade development strategy of China has been to expand imports, and to keep the balanced development of import and export [19]. Scholars have demonstrated the economic benefits of lowing tariffs. Based on the tariff reduction schedules of FTA negotiations between China and South Korea, for example, some studies have found that bilateral tariff reductions would promote trade creation, increasing the GDP, boosting import and export and raising welfare by means of the calculation of GTAP model [20, 21]. Furthermore, lowering import tariffs would not only help raising the income, consumption, employment, social welfare, return on capital and output level, but also reduce price level and enhance economic benefits for trade partners. Obviously, imposing tariffs on goods with trade complementarity is not in line with bilateral economic interests in the long-term and decreases social welfare for both sides. It is necessary for jointing efforts to decrease tariff, so that speed up the negotiation process of free trade agreements between China and Norway.

<sup>2</sup> Source: Tariff HS classification data of the WTO from 2013 to 2017.

#### 4. Conclusions and suggestions

Promoting tariff relief for commodities with strong trade complementarity between China and Norway should be taken into account first. The strategies to push up free trade agreement should be step by step.

First, the key of pushing up the negotiation of the free trade agreement between both sides should focus on cutting down tariff on Chinese SITC 6. Chinese side needs to strive for cutting tariff to zero for some overcapacity goods within SITC 6, including textiles, steel products, building materials, non-ferrous metals and glass products, etc., if it is difficult for Norway to reduce the tariff on various commodities of SITC 6 to zero at one time. The tariff on the rest, such as leather products, ceramic products, paper and paperboard products, could be reduced by stages. Concurrently, the increase of investment and consumption demand for SITC 6 in Norway will be benefit from lowering tariff on Chinese SITC 6.

Second, it is important for China to reduce the tariff of 7.1% on Norwegian SITC 2 and 12.4% on SITC 4 as priority through signing tariff reciprocity agreement or arranging schedule for tariff reduction. Since SITC 2 and SITC 4 are group commodities with strong trade complementarity with China, of which some goods, such as iron, copper, nickel and aluminum ores, concentrates, raw fur, rubber, oilseeds, wood corks, pulp waste paper, silk, cotton, hemp textile fibers and fats and waxes of animal and vegetable, are all shortage resources in urgent need in China. Reducing the tariff on Norwegian SITC 2 will help to stabilize and to expand the production chain of raw materials and finished products because the group SITC 2 is the upstream products of SITC 6 of China.

Third, SITC 3 of Norway is highly complementary to China and has great trade potentiality. If Chinese tariff on SITC 3 is further reduced on the basis of improving transportation infrastructure, it will greatly stimulate the volume of bilateral trade.

Finally, lowering the tariff of 10.3% on Norwegian SITC 0 could alleviate the shortage of meat food in China. Although the current TCI of SITC 0 of Norway is less than 1, according to the trend of Chinese meat consumption, the demand for aquatic products of Norway will continue to rise, and the potentiality of supply and demand complementarity of SITC 0 between the two countries is extremely huge.

For the sake of reaching the FTA, both sides should gradually implement zero tariff on trade complementary commodities with strong domestic demand through gradually decreasing the overall tariff level, which will not only help stabilize and improve Chinese position in the international industrial chain and supply chain, but also help to achieve the cooperation of an all-win result between China and Norway.

#### References

- [1] Li G, Ye X. Discussion on the rationality of Chinese tariff level and tariff structure under the new situation. *International Trade Issues*. 2017(7): 23-35.
- [2] Salvador DE. *International Economics*. Beijing: Tsinghua University Press; 2011. P. 57.
- [3] Robert J, Porter M, Redding S. Specialization dynamics. *Journal of International Economics*. 2014(58):299–334.
- [4] Stack MM, Bliss M. EU economic integration agreements, brexit and trade. *Review of World Economics*. 2020(5):1-31.  
Li G, Ye X. Discussion on the rationality of Chinese tariff level and tariff structure under the new situation. *International Trade Issues*. 2017(7): 23-35.
- [5] Zheng ZY, Meng M. Analysis of the economic effects of the Asia Pacific Free Trade Area. *International*

- Economic Cooperation. 2017(7): 51-65.
- [6] Andreosso-O', Callaghan B. Economic structural complementarity: how viable is the Korea-EU FTA? Journal of Economic Studies. 2009(36):147–167.
- [7] Hu Y, Yan JL, Quan Y. Empirical Study on the measurement of trade complementarity between China and countries along the “21st century Maritime Silk Road” and its influencing factors. World Economic Research. 2017(8): 39-50.
- [8] Liu CP, Xiao HF. China's agricultural trade with North European countries – complementarity, competitiveness and development potential under “one belt and road”. Journal of Dalian University of Technology. 2019 (4):67-77.
- [9] Chen YH, Lai MY. Analysis on trade structure and competitive complementarity between China and Norway. Financial Theory and Practice. 2011(9): 52-65.
- [10] Bojnec S, Ferto I. Complementarities of trade advantage and trade competitiveness measures. Applied Economics. 2012(44):399 – 408.
- [11] Chen M, Chen SM. The study of potential trade effects of interconnection and interconnection on China under “one belt and road”. Modern Economy Exploration. 2019(12): 59-70.
- [12] Hofmann C, Alberto O, Michele R, Horizontal D. A new database on the content of deep agreements. Policy Research Working Paper of NBER. 2017(5): 7981-7997.
- [13] Zhou XB, Chen Z. The strategic connotation and economic impact of actively expanding imports. Modern Economy Exploration. 2020(2): 17-29.
- [14] Aditya A, Acharyya R. Trade liberalization and export diversification. International Review of Economics and Finance. 2015(39):390-410.
- [15] Ministry of Climate and Environment. Report to the Storting. Norway. 2015-2016: 1-120.
- [16] Yi XL. Bright development prospects of China and the future cooperation between China and Norway. Speech at the annual forum of Chamber of Commerce both Norway and China; Embassy of P. R. China in the Kingdom of Norway, 2021 March 26, Oslo, Norway.
- [17] Thured E. Reaching a free trade agreement with China is the top priority of the Norwegian government. Embassy of P. R. China in the Kingdom of Norway. 2018 Aug 23, Oslo. Available online: <http://no.mofcom.gov.cn> (accessed on 2019 Dec. 26).
- [18] Jensen S. Agreement with China to renegotiate tax treaty. Press release. Norwegian government. 2017 Sept. 28, Oslo. Available online: <http://www.regjeringen.no> (accessed on 2020 Jan. 22).
- [19] Zhu QR, Yang L, Wang Y. Evaluation of economic benefits of China's import tariff reduction. World Economy Studies. 2021(2):19-32.
- [20] Zhou SD, Xiao X, Yang J. Analysis of the economic influence of the establishment of the China Korea FTA on their major industries based on the tariff reduction schedules of the bilateral FTA negotiation. Issues of International Trade. 2016(5): 116-129.
- [21] Lawrence R. World tariff rate declines and progress in global trade liberalization: multilateral, regional and unilateral. Research on Financial Development. 2021(7): 42-47.

# The SAPEVO-M-NC Method

Sérgio Mitihiro do Nascimento MAÊDA<sup>a,d,1</sup>, Marcio Pereira BASÍLIO<sup>b</sup>, Igor Pinheiro de Araújo COSTA<sup>a,d</sup>, Miguel Ângelo Lellis MOREIRA<sup>a,d</sup>, Marcos dos SANTOS<sup>c,d</sup>, and Carlos Francisco Simões GOMES<sup>d</sup>

<sup>a</sup>Federal Fluminense University, Brazil

<sup>b</sup>Military Police of the State of Rio de Janeiro, Brazil

<sup>c</sup>Military Institute of Engineering, Brazil

<sup>d</sup>Naval Systems Analysis Center, Brazil

**Abstract.** This article aims to present with more details, the multicriteria decision aid SAPEVO-M-NC (Simple Aggregation of Preferences Expressed by Ordinal Vectors - Non-Compensatory - Multi Decision Makers). It is a new version of the SAPEVO-M method, of an ordinal, non-compensatory nature and with the possibility of acting by multiple decision makers. As a result, the method provides information on the partial weights, indicating the relative importance of the criteria for each of the decision makers, the relative dominance values and two evaluations on the performance of the alternatives: a partial one, which considers the absolute dominance indices, being used to assess existing dominance relationships; and a global one, which provides the performance rates of the alternatives, making it possible to order them as well as to carry out a sensitivity analysis on the observed performances, reflecting in greater transparency in the decision-making process.

**Keywords.** Ordinal ranking methods; Multicriteria decision aid; Non-compensatory.

## 1. Introduction

Fundamentally, we are all decision makers. Everything we do, consciously or not, is the result of a decision-making process, involving various information that must be evaluated to provide greater transparency to the process and understanding of the system by the decision maker(s) [1][2][3].

The ordinal methods were the first methods to aid decision-making, developed after the mid-eighteenth century, by studies by Jean-Charles de Borda [4]. These methods offer advantages due to their relative ease of understanding and administration, as well as greater reliability in terms of reducing inconsistencies generated in the preference elicitation process[5]. Due to the need to require less cognitive effort from the decision maker, it is expected that they reflect their preferences more accurately [4][6][7]. This paper presents the method called SAPEVO-M-NC [8]. The method allows to aggregate, through an ordinal process, the preferences of decision makers regarding the importance of the criteria and the performance of the alternatives.

---

<sup>1</sup> Corresponding Author, Sérgio Mitihiro do Nascimento Maêda, Federal Fluminense University - UFF, 24210-240, Niterói, Brazil; E-mail: [sergiommaeda@gmail.com](mailto:sergiommaeda@gmail.com).

This research was financially supported by Military Police of the State of Rio de Janeiro.

Although robust, it uses a relatively simple axiomatic base, resulting in low cognitive effort on the part of evaluators.

## 2. The SAPEVO-M-NC Method

Derived from the SAPEVO-M method [4], the SAPEVO-M-NC method [8], consists of an ordinal method, non-compensatory nature, which aims at the problem of ordering ( $P\gamma$ ), and with the possibility of acting by multiple decision makers. In the method, the evaluation of the performance of the alternatives is carried out directly, with no need to carry out parity comparisons between the alternatives to obtain the modeling of preferences, resulting in a substantial reduction in the cognitive effort on the part of the Decision Makers DMs. The method allows for two assessments, one partial and one global, resulting in a more sensitive analysis of the performance of alternatives, as well as greater transparency about the decision-making process [8].

Unlike Classic Decision Theory, which basically considers two supposedly transitive preference relations, designated by Indifference (I) and by Strict Preference (P), this method is based on the Fundamental System of Preference Relations (FSPR), also incorporating the weak preference relationship (Q). According to [9], there are several reasons why researchers seek to avoid the type of modeling that is based on the axiom of complete comparability and transitivity between alternatives, among which, can be mentioned the fact that the decision maker does not have all the information that allows him to choose one of the alternatives and by forcing a relationship of strict preference, or indifference, could lead to arbitrary and inconsistent errors.

After the DMs establish the criteria and alternatives, the method can be divided into five steps:

Step 1 - Ordinal transformation of the preferences of each DM, in each criterion, which are added at the end of this step, giving rise to a vector ( $V_i$ ), representing the weights of the criteria. Table 1 shows the relationship of relative importance between the criteria:

**Table 1.** Relative importance between the criteria

Scale 1	Verbal representation	Scale 2
$\delta(c_i c_j) \ll 1 \leftrightarrow c_i <<< c_j$	Absolutely less important	-3
$\delta(c_i c_j) \preccurlyeq 1 \leftrightarrow c_i << c_j$	Much less important	-2
$\delta(c_i c_j) \prec 1 \leftrightarrow c_i < c_j$	Less important	-1
$\delta(c_i c_j) = 1 \leftrightarrow c_i \approx c_j$	Equally important	0
$\delta(c_i c_j) \succ 1 \leftrightarrow c_i > c_j$	More important	1
$\delta(c_i c_j) \succcurlyeq 1 \leftrightarrow c_i >> c_j$	Much more important	2
$\delta(c_i c_j) \gg 1 \leftrightarrow c_i >>> c_j$	Absolutely more important	3

Let  $DM_k$  be a set of "k" decision-makers,  $D = \{DM_1, DM_2, \dots, DM_k\}$  that express their opinions on the relative importance of the criteria involved. These preferences give rise to the  $MDM_k$  preference matrix. The relationship between the two scales of the table allows the transformation of the matrix (1) into (2):

$$MDM_k = [\delta(c_i|c_j)], \text{ in a column vector } [V_i], \text{ where:} \quad (1)$$

$$V_i = \sum_{j=1}^m \delta(c_i|c_j) \quad (i = 1, \dots, m, \text{ and } k = 1, \dots, n) \quad (2)$$

After generating the vector  $V_i$ , its  $a_{ij}$  elements are normalized according to (3):

$$v = (a_{ij} - \min a_{ij}) / (\max a_{ij} - \min a_{ij}) \quad (3)$$

Giving rise to the  $DM_k$  preferences vector. If null values occur in this step, they are replaced by 1% of the second lowest value obtained. After all DM's carry out their evaluations, the normalized vectors are added, giving rise to the weight vector that expresses the importance of the criteria [4].

Step 2 - Ordinal classification ( $\Theta_{ij}$ ) of the performance of the alternatives:

In this step, each DM assigns the ratings related to the performance of the alternatives in each criterion (table 2), which are related to their rating ranges  $g_{(ij)}$ . After all "k" DMs perform their evaluations, the arithmetic mean  $\mu_{(ij)}$  of the classification ranges of the performances of the alternatives in each criterion is obtained.

**Table 2.** Ordinal ratings of performance of alternatives

Ordinal classification ( $\Theta_{ij}$ ) of the performance of alternative $i$ in criterion $j$	Classification range $g_{(ij)}$
Excellent (E)	1
Very Good (VG)	2
Good (G)	3
Medium (M)	4
Bad (B)	5
Very Bad (VB)	6
Poor (P)	7

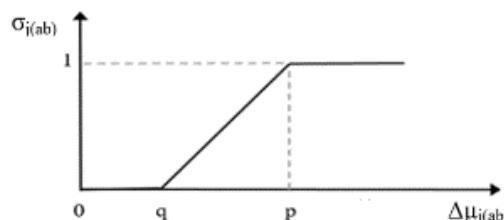
Step 3 – Obtaining the fractions of the criteria weights ( $\sigma_{j(ab)}$ ).

For each criterion "j", a parity comparison is made between the alternatives to verify the relative distance between the mean values of the classification ranges (4):

$$\Delta\mu_{i(ab)} = \mu_{(ia)} - \mu_{(ib)} \quad (4)$$

This value allows identifying in the preference modeling (figure 1 and table 3) the weight fraction of criterion "j", obtained by alternative "a" in relation to alternative "b" ( $\sigma_{j(ab)}$ ).

**Figure 1.** Preference function of a criterion with linear variation.



**Table 3.** Criteria preference modeling

Indifference (I)	$\mu_{(ia)} - \mu_{(ib)} \leq 1 : \sigma_{(ab)} \rightarrow 0$
------------------	--

Weak Preference (Q)	$1 < \mu_{(ia)} - \mu_{(ib)} \leq 3$	$\rightarrow \frac{(A_{ij} - \min A_{ij})}{(\max A_{ij} - \min A_{ij})}$
Strong Preference (P)	$3 < \mu_{(ia)} - \mu_{(ib)}$	$\therefore \sigma_{(ab)} \rightarrow 1$

Step 4 – Calculation of relative  $d_{ab}$  dominance.

Obtained by the weighted sum of the criteria weights ( $w_j$ ), with the corresponding fraction ( $\sigma_{j(ab)}$ ) verified in the preference modeling (5):

$$d_{ab} = \sum w_j \times \sigma_{j(ab)} \quad (5)$$

Step 5 - Conducting assessments

5.1 - Partial evaluation: Calculation of Absolute  $D_{ab}$  Dominance and Outranking Rate  $\eta_{ab}$ . The difference between the relative dominances  $d_{ab} - d_{ba}$  provides information on the absolute dominance  $D_{ab}$  between the alternatives (6), where positive values of " $D_{ab}$ ", indicate that alternative "a" dominated alternative "b".

$$D_{ab} = d_{ab} - d_{ba} \quad (6)$$

This information allows to identify the existing relationships between the alternatives, and the assembly of a dominance graph. Dividing  $D_{ab}$ , by the sum of the weights, the percentage rate of absolute dominance is obtained (7).

$$\eta_{ab} = D_{ab} / (\sum w_j) \quad (7)$$

This information allows the DM more clarity about the partial performances between the alternatives.

5.2 - Global assessment: The method makes it possible to carry out an analysis of the total performance of each of the "n" alternatives, evaluating the Performance Rates ( $T_a$ ) obtained by each one (8):

$$T_a = \sum d_{ab} / (\sum w_j \times (n-1)) \quad (8)$$

This information allows greater transparency of the process to the DM, especially in situations where:  $D_{ab} = D_{ba} = 0$ . Finally, the results allow the ordering of alternatives.

### 2.1. Application of SAPEVO-M-NC (Numerical example)

To elucidate the steps presented, consider for example composed of seven alternatives ( $A_1, A_2, A_3, A_4, A_5, A_6, A_7$ ), which are compared considering four criteria ( $C_1, C_2, C_3, C_4$ ), by three DM's = ( $DM_1, DM_2, DM_3$ ). Table 4 shows as an example the ordinal evaluation of the criteria, by  $DM_1$ :

**Table 4.** Example of ordinal assessment of the importance of attributes by  $DM_1$

	$C_1$	$C_2$	$C_3$	$C_4$
$C_1$	Equally important	More important	Absolutely more important	Much more important
$C_2$	Less important	Equally important	Much more important	More important

$C_3$	Absolutely less important	Much less important	Equally important	Less important
$C_4$	Much less important	Less important	More important	Equally important

The table 5 presents the values of scale 2 related to the ordinal evaluations, together with the normalized weight vector.

**Table 5.** Criteria preference modeling

	$C_1$	$C_2$	$C_3$	$C_4$	$V_i$	Normalized vector
$C_1$	0	+1	+3	+2	6	1
$C_2$	-1	0	+2	+1	2	0,666667
$C_3$	-3	-2	0	-1	-6	0,003333
$C_4$	-2	-1	+1	0	-2	0,333333

The table 6 presents the evaluation of the importance of the criteria by all DM's.

**Table 6.** Evaluation of the importance of the criteria by the DMs

	DM <sub>1</sub>	DM <sub>2</sub>	DM <sub>3</sub>	Final weights
$C_1$	1	0,875	1	2,875
$C_2$	0,6667	1	0,5555	2,222
$C_3$	0,0033	0,00875	0,2222	0,234
$C_4$	0,3333	0,125	0,0022	0,461

After defining the weights of the criteria, the next step is to classify the performance of the alternatives by the DMs. The table 7 presents the performance ratings of the alternatives with the corresponding cardinal values:

**Table 7.** Evaluation of the importance of the criteria by the DMs

DM <sub>1</sub>				DM <sub>2</sub>				DM <sub>3</sub>				Arithmetic mean $\mu_{(ij)}$							
$C_1$	$C_2$	$C_3$	$C_4$	$C_1$	$C_2$	$C_3$	$C_4$	$C_1$	$C_2$	$C_3$	$C_4$	$C_1$	$C_2$	$C_3$	$C_4$				
A <sub>1</sub>	1	2	1	1	A <sub>1</sub>	1	3	1	1	A <sub>1</sub>	1	2	1	1	A <sub>1</sub>	1	2,33	1	1
A <sub>2</sub>	6	5	3	3	A <sub>2</sub>	6	6	3	3	A <sub>2</sub>	6	6	3	3	A <sub>2</sub>	6	5,66	3	3
A <sub>3</sub>	6	4	6	5	A <sub>3</sub>	6	5	6	5	A <sub>3</sub>	6	5	5	6	A <sub>3</sub>	6	4,66	5,66	5,33
A <sub>4</sub>	3	1	2	1	A <sub>4</sub>	2	1	2	1	A <sub>4</sub>	2	1	2	1	A <sub>4</sub>	2,33	1	2	1
A <sub>5</sub>	5	5	6	2	A <sub>5</sub>	5	6	6	2	A <sub>5</sub>	4	6	5	2	A <sub>5</sub>	4,66	5,66	5,66	2
A <sub>6</sub>	3	4	6	2	A <sub>6</sub>	2	5	5	2	A <sub>6</sub>	2	5	5	2	A <sub>6</sub>	2,33	4,66	5,33	2
A <sub>7</sub>	4	3	3	4	A <sub>7</sub>	4	4	3	4	A <sub>7</sub>	3	4	3	4	A <sub>7</sub>	3,66	3,66	3	4

Table 8 presents, respectively, the difference between the performance evaluations of the alternatives  $\Delta\mu_{j(ab)}$ , the values obtained in the preference modeling  $\sigma_{j(ab)}$ , the calculation of the relative dominance  $d_{ab}$ , the absolute dominance  $D_{ab}$  and the outranking rate  $\eta_{ab}$ .

**Table 8.** Values of  $\Delta\mu_{j(ab)}$ ,  $\sigma_{j(ab)}$ ,  $d_{ab}$ ,  $D_{ab}$  e  $\eta_{ab}$

$\Delta\mu_{j(ab)} = \mu_{j(a)} - \mu_{j(b)}$				$\sigma_{j(ab)}$				$d_{ab} = \sum w_j \times \sigma_{j(ab)}$				$D_{ab}$	$\eta_{ab}$	
	$C_1$	$C_2$	$C_3$	$C_1$	$C_2$	$C_3$	$C_4$	$C_1$	$C_2$	$C_3$	$C_4$	Total	Total	
A <sub>1</sub> -A <sub>2</sub>	5,00	3,33	2,00	2,00	1,00	1,00	0,50	0,50	2,88	2,22	0,12	0,23	5,44	0,94
A <sub>1</sub> -A <sub>3</sub>	5,00	2,33	4,67	4,33	1,00	0,67	1,00	1,00	2,88	1,48	0,23	0,46	5,05	0,87
A <sub>1</sub> -A <sub>4</sub>	1,33	-1,33	1,00	0,00	0,17	0,00	0,00	0,00	0,48	0,00	0,00	0,00	0,48	0,11
A <sub>1</sub> -A <sub>5</sub>	3,67	3,33	4,67	1,00	1,00	1,00	1,00	0,00	2,88	2,22	0,23	0,00	5,33	0,92
A <sub>1</sub> -A <sub>6</sub>	1,33	2,33	4,33	1,00	0,17	0,67	1,00	0,00	0,48	1,48	0,23	0,00	2,19	0,38
A <sub>1</sub> -A <sub>7</sub>	2,67	1,33	2,00	3,00	0,83	0,00	0,50	1,00	2,40	0,00	0,12	0,46	2,97	0,51

A <sub>2</sub> -A <sub>3</sub>	0,00	-1,00	2,67	2,33	0,00	0,83	0,67	0,00	0,00	0,20	0,31	0,50	0,50	0,09
A <sub>2</sub> -A <sub>4</sub>	-3,67	-4,67	-1,00	-2,00				0,00	0,00	0,00	0,00	0,00	-5,33	-0,92
A <sub>2</sub> -A <sub>5</sub>	-1,33	0,00	2,67	-1,00		0,00	0,83		0,00	0,00	0,20	0,00	-0,28	-0,05
A <sub>2</sub> -A <sub>6</sub>	-3,67	-1,00	2,33	-1,00			0,67		0,00	0,00	0,16	0,00	-2,72	-0,47
A <sub>2</sub> -A <sub>7</sub>	-2,33	-2,00	0,00	1,00			0,00	0,00	0,00	0,00	0,00	0,00	-3,03	-0,52
A <sub>3</sub> -A <sub>4</sub>	-3,67	-3,67	-3,67	-4,33				0,00	0,00	0,00	0,00	0,00	-5,79	-1,00
A <sub>3</sub> -A <sub>5</sub>	-1,33	1,00	0,00	-3,33		0,00	0,00		0,00	0,00	0,00	0,00	-0,46	-0,08
A <sub>3</sub> -A <sub>6</sub>	-3,67	0,00	-0,33	-3,33		0,00			0,00	0,00	0,00	0,00	-3,34	-0,58
A <sub>3</sub> -A <sub>7</sub>	-2,33	-1,00	-2,67	-1,33				0,00	0,00	0,00	0,00	0,00	-2,19	-0,38
A <sub>4</sub> -A <sub>5</sub>	2,33	4,67	3,67	1,00	0,67	1,00	1,00	0,00	1,92	2,22	0,23	0,00	4,37	0,75
A <sub>4</sub> -A <sub>6</sub>	0,00	3,67	3,33	1,00		0,00	1,00	1,00	0,00	0,22	0,23	0,00	2,46	0,42
A <sub>4</sub> -A <sub>7</sub>	1,33	2,67	1,00	3,00	0,17	0,83	0,00	1,00	0,48	1,85	0,00	0,46	2,79	0,48
A <sub>5</sub> -A <sub>6</sub>	-2,33	-1,00	-0,33	0,00				0,00	0,00	0,00	0,00	0,00	-1,92	-0,33
A <sub>5</sub> -A <sub>7</sub>	-1,00	-2,00	-2,67	2,00				0,00	0,00	0,00	0,23	0,23	-1,08	-0,19
A <sub>6</sub> -A <sub>7</sub>	1,33	-1,00	-2,33	2,00	0,17		0,50	0,50	0,48	0,00	0,00	0,23	0,71	0,55

The positive values for  $D_{ab}$ , indicates that the first alternative dominates the second one of the analyzed pair. Through the results obtained, we verified the following overcoming relationship:  $A_1 > A_4 > A_6 > A_7 > A_5 > A_2 > A_3$ .

And the graph with the dominance relationships can be constructed (figure 2):

**Figure 2.** Graph representing the dominance relationships between the alternatives.



It is observed that with the  $D_{ab}$  values, it was possible to establish all the dominance relationships and order the alternatives, however, for a deeper analysis, the performance rates ( $T_a$ ) were obtained by each alternative.

$$A_1 = 0,6179; A_4 = 0,6074; A_6 = 0,2542; A_7 = 0,1921; A_5 = 0,0336; A_2 = 0,0245; A_3 = 0,0000$$

### 3. Analysis of results and Conclusion

Ordinal methods have a wide field of application and, due to their nature, are closer to the way people make their decisions in the face of processes that deal with qualitative variables, partial or incomplete information.

Through the numerical example presented, it was verified that the SAPEVO-M-NC method: allowed to aggregate, through an ordinal process, the preferences of the decision makers regarding the importance of the criteria and the performance of the alternatives; although robust, it uses a relatively simple axiomatic base; the fact of not exploring parity comparisons between alternatives to perform preference modeling results in less cognitive effort on the part of evaluators; the sensitivity analysis on the performance of the alternatives allows for a more in-depth assessment, resulting in greater transparency to the decision maker(s) about the decision-making process developed.

Due to the relative ease of application, associated with a low cognitive effort on the part of the evaluators, it is concluded that this methodology can provide great gains, not only for the academic community, but also for society as a whole, presenting itself

as an alternative tool for multicriteria decision support, of an ordinal, non-compensatory nature and with the possibility of supporting multiple decision makers.

As a proposal for future work, this method can be better explored by being approached in case studies comparing the results with those obtained through other methods in the literature.

## References

- [1] Saaty TL. Decision making with the analytic hierarchy process. *Int J Serv Sci* [Internet]. 2008;1(1):83. Available at: <http://www.inderscience.com/link.php?id=17590>
- [2] Moreira MÂL, Costa IP de A, Pereira MT, dos Santos M, Gomes CFS, Muradas FM. PROMETHEE-SAPEVO-M1 a Hybrid Approach Based on Ordinal and Cardinal Inputs: Multi-Criteria Evaluation of Helicopters to Support Brazilian Navy Operations. *Algorithms* [Internet]. 2021;14(5):140. Available at: <https://www.mdpi.com/1999-4893/14/5/140>, DOI: 10.3390/a14050140
- [3] Santos M dos, Costa IP de A, Gomes CFS. Multicriteria Decision-Making In The Selection Of Warships: A New Approach To The Ahp Method. *Int J Anal Hierarchy Process*. 2021;13(1):147–6, DOI: 10.13033/ijahp.v13i1.833
- [4] Gomes CFS, Santos M dos, Teixeira LFH de S de B, Sanseverino AM, Barcelos MR dos S. SAPEVO-M: A group multicriteria ordinal ranking method. *Pesqui Operacional* [Internet]. 2020;40, DOI: 10.1590/0101-7438.2020.040.00226524
- [5] Moreira MÂL, Gomes CFS, dos Santos M, do Carmo Silva M, Araujo JVGA. PROMETHEE-SAPEVO-M1 a Hybrid Modeling Proposal: Multicriteria Evaluation of Drones for Use in Naval Warfare. In: International Joint conference on Industrial Engineering and Operations Management [Internet]. Military Institute of Engineering (IME), Urca, RJ 22290-270, Brazil: Springer; 2020. p. 381–93. Available at: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85097128127>, DOI: 10.1007/978-3-030-56920-4\_31
- [6] Rodrigues MVG, Duarte TE, dos Santos M, Gomes CFS. Prospective scenarios analysis: Impact on demand for oil and its derivatives after the COVID-19 pandemic. *Brazilian J Oper Prod Manag*. 2021;18(2), DOI: 10.14488/BJOPM.2021.039.
- [7] Costa IP de A, Sanseverino AM, Barcelos MR dos S, Belderrain MCN, Gomes CFS, dos Santos M. Choosing flying hospitals in the fight against the COVID-19 pandemic: structuring and modeling a complex problem using the VFT and ELECTRE-MOr methods. *IEEE Lat Am Trans*. 2021;19(6):1099–106, DOI: [10.1109/TLA.2021.9451257](https://doi.org/10.1109/TLA.2021.9451257)
- [8] Maêda SM do N, Costa IP de A, Santos M dos., Gomes CFS. Economic and edaphoclimatic evaluation of Brazilian regions for African mahogany planting - an approach using the SAPEVO-M-NC ordinal method. The International Conference on Information Technology and Quantitative Management (ITQM), organizador. Chengdu: The International Conference on Information Technology and Quantitative Management (ITQM), Elsevier - Procedia Computer ScienceProcedia Computer Science; 2021.
- [9] GOMES LFAM, GOMES CFS. Princípios e métodos para a tomada de decisão: Enfoque multicritério. 6º. São Paulo: Atlas; 2019.

# The Support System for Anomaly Detection with Application in Mainframe Management Process

Dominik STRZAŁKA<sup>a</sup>, Alicja GERKA<sup>a,1</sup>, Bartosz KOWAL<sup>a</sup>, Paweł KURASŁA<sup>a</sup>,

Grzegorz LEOPOLD<sup>b</sup>, Michał LEWICZ<sup>b</sup> and Dawid JAWORSKI<sup>c</sup>

<sup>a</sup>*Department of Complex Systems, Rzeszów University of Technology,*

*Al. Powstańców Warszawy 12, 35-959 Rzeszów, POLAND*

<sup>b</sup>*Z-RAYS, Plac Andersa 7, 61-894 Poznań, POLAND*

<sup>c</sup>*Department of Mathematical Modelling, Rzeszów University of Technology,*

*Al. Powstańców Warszawy 12, 35-959 Rzeszów, POLAND*

**Abstract.** The process of mainframe machines managing and administration requires not only specialized expert knowledge based on many years of experience but also on appropriate tools provided by a machine performance management system, e.g. the Resource Measurement Facility (RMF). The aim of this paper is to show some preliminary results of Z-RAYS system construction that is built basing on machine learning (ML) techniques. It allows automatic detection of anomalies and generation of early warnings about some errors that can appear in the mainframe to support mainframe management process. Presented results are based on extensive simulations that were done basing on the IBM emulator. We focus on determining the degree of the metrics variability, the degree of the data repeatability in metrics, some approaches in metrics anomaly detection and solutions for event correlation detection in metrics.

**Keywords.** Mainframe, anomaly detection, support system, machine learning, Z-RAYS

## 1. Introduction

Mainframe (mainframe machines, big iron computers) are a class of computers used mainly by large organizations for critical applications like financial, statistical. Today, this term (name) is usually related to computers compatible with the IBM System/360 line introduced in 1965 [1]. There is no formal definition of a mainframe. Such machines run uninterrupted for very long periods of time and are used everywhere when the high availability is required because any downtime would be costly or even catastrophic. They can run in parallel different instances of operating systems thanks to the technique of virtual machines. They are not a supercomputer [1].

In this paper we focus on a series of preliminary studies and analyses to prepare:

---

<sup>1</sup> Corresponding Author, Alicja Gerka, Department of Complex System, Rzeszów University of Technology, Al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland; E-mail: a.gerka@prz.edu.pl.

- recommendations for metrics,
- proposal of suggested algorithms and possible approaches to anomaly detection,
- preparation and presentation of the results of the performed series of experiments,
- preliminary studies on correlating events that are anomalies.

The aim of this paper is to solve three main problems and show some preliminary solution results. The first problem to be solved is the development of a methodology for detecting potential anomalies in time series representing the metrics. One possible approach to solve this issue is to only refer to statistical analyzes, but another approach is to rely on ML algorithms related to time-series analyses. Bearing in mind the fact that from a mainframe machine more than 15 thousand RMF metrics (but not only) can be constantly analyzed, an approach to automatize this process is expected, in particular making decisions on how the data from the metric will be processed. The second aspect is to identify one or more approaches that allow the use of different anomaly detection methods together with preliminary research of these solutions. This approach is about the proper detection of anomalies, also with the indication of the moment when such an anomaly value appears or there is a significant deviation from the adopted metric pattern behavior. The third element is related to the choose of solution for correlating events (anomalies in time series) and the detection of hidden dependencies (or correlations) between these anomalies.

The research and development problem is: how, for collected, visualized, and analyzed data, to create a support system that allows to take the right administration decisions and ensure uninterrupted mainframe work.

## 2. Preliminaries and methods

The source of all provided metrics is a special mainframe emulator (IBM Z Development and Test Environment) and RMF, a mainframe performance management tool that measures selected areas of system activity and presents the collected data in the form of metric records [2].

In our carried study, the supplied dataset contained 42 metric datasets generated on a mainframe emulator with the following unique SOURCEID's like 8D30F0, 8D0360, 8D1040 which showed selected grouped metrics. Each such metric can consist of many time series and as part of each set of metrics (SOURCEID), data on various types of devices (DEV) and processes (RCNAME) are presented via time series with different time resolutions. From the first raw data, it was possible to generate over 500 unique metrics in the form of time series containing numerical data with a time stamp.

The provided metrics under the name RPNAME (RePort Name) are very extensive, containing many columns of various data, but from our point of view, the most important columns are VALUE and dateAndTime. Other relevant columns may indicate: DEV - external device name, SYSTEM - system name, RCNAME - MASTER, OMVS and their correct interpretation leads to the final metric suitable for analysis in the form of a time series with VALUE and dateAndTime.

We decided to take the resolution of 1 minute for each stored data. Originally, the data was saved as a text file, but for large systems, it could mean the need to collect up to more than 500 GB of historical data per year. Using a different recording format, e.g.

binary, assuming a regular (every minute) recording only the metric values in a bit format (4 bytes per number) were stored, and in the absence of a byte record the value -1 was signed. The total amount of data was significantly limited to 5,760 bytes per day per metric. It should be noted here that many RPNAME metrics (Table 1) contain many data relating to different DEVs within the same timestamp (so-called splitting).

**Table 1.** Some of the original set of RPNAME metrics for the studies

#	SOURCEID	RPNAME
1	8D16D0	delayforXCFbyWLMreportclass
2	8D0E90	Ioactivityrate
3	8D12A0	Iointensitybyvolume
4	8D12C0	IOSqueuetimebyvolume
5	8D27B0	MSUbyWLMreportclass

### 2.1. Determining the degree of the metrics variability

After the first month of data gathering, graphical visualizations of metrics were done. They showed that some of them are constant (do not change during the time) or contain only two values, -1 and 0. Mainframe experts decided that this could be a normal behavior related for instance to the fact, that some processes in the mainframe machine can be visible only for short periods, several times per year, or when the system is in a normal state. We decided to classify all metrics to select those, which have high volatility. Among the numerical characteristics of doing such classification, it should be mentioned for example: analysis of the histogram, searching for a modal value, and testing the power spectrum to determine periods (days, weeks) of repeating certain processes in the metrics. This problem is critical to the proposed ML algorithms.

One of the proposed ways to determine the degree of variability (1) of the metric may be to examine the process increments:

$$D_V(t) = Y(t+1) - Y(t), \quad (1)$$

where:  $Y(t)$  is the value of metrics in  $t$  – time,  $Y(t+1)$  is the next value of metrics in  $t$  – time.

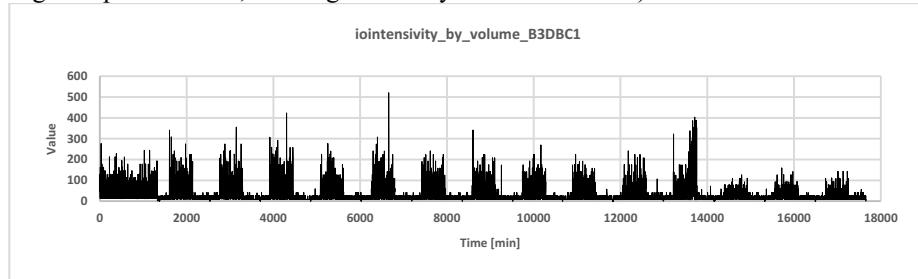
In this way, each increase/decrease in the value of the metric will be visible in the new time series and the periods when the metric has constant values (not necessarily = 0) will be treated as no variation. Then, for a new time series of process increments, the modal value can be determined. Keeping in mind that the system was observed for 7 days, 24 hours a day with a resolution of 1 minute, it generates  $60 * 24 * 7 = 10,080$  data per metric (time series) during one week. For example, if for more than 99.9% of cases the modal value will be 0, then the time series has a very low variation, 99.9% of cases equal to 0 give ~10 changes in the value of the metric during the week and ~1.5 per day. It was called Category A. Details are in Table 2. This approach can be changed: more categories can be added or different values for reference points determined.

**Table 2.** Proposed degree of the metrics variability

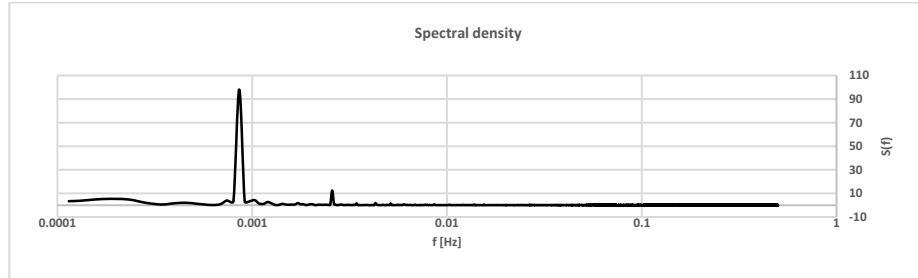
Modal value $x$ for 0	Category	Changes of values in metrics
More than 99.9%	A	< 1.5 per day
$99.9\% \leq x < 99\%$	B	< 15 per day
$99\% \leq x < 90\%$	C	< 150 per day
Other	D	more than 150

## 2.2. Determining the degree of the data repeatability in metrics

The visual analysis for selected metrics indicates a certain cyclicity of the processes; determining the period of this cyclicity may be of key importance for selecting the parameters of ML algorithms. One way to detect and quantify this cyclicity is the (power) spectrum of the metrics. The occurrence of distinct power peaks for specific frequencies in the power spectrum will suggest that some characteristic processes of a specific period (inverse of the frequency) are occurring. This period can be expressed in hours, days, weeks. The fast Fourier transform was used [3] and some results of this approach are visible in Figs. 1 and 2, where the frequency axis (Fig. 2) is given on a log scale (the lowest values of frequency are the most important because they indicate the longest repetition time; on a log scale they are better visible).



**Figure 1.** Time plot for the iointensitybyvolume-B3DBC1 metric for 15 days.



**Figure 2.** Spectrum  $S(f)$  for the iointensitybyvolume-B3DBC1 metric with one visible dominant frequency.

In the used method one can calculate that the number of process repetitions is equal to the number of collected observations multiply by the dominant frequency  $f$ , where

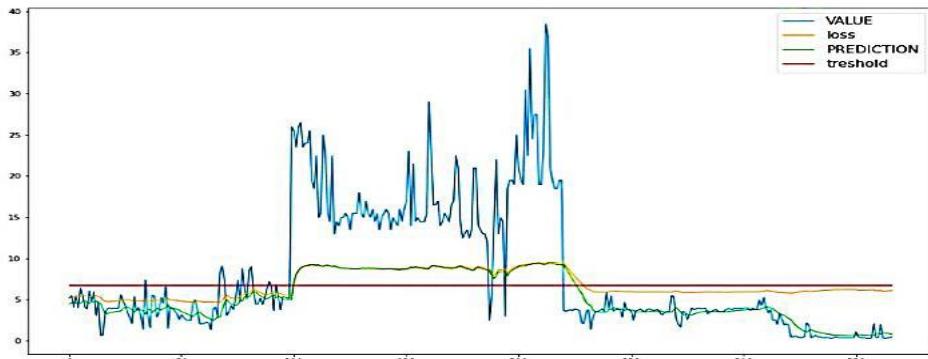
$f \in <0, 0.5>$ . Having in mind that during one day there are 1440 minutes we can see that: the number of collected data/1440/the number of repetitions determines how many days the process is repeated. This defines the TIME\_STEPS parameter in ML algorithm. In Fig. 2 there is one distinct peak for  $f = 0.000857339$  Hz. Given that 17498 input data of the time series were analyzed, it can be calculated that  $17498 * 0.000857339 = 15$ . This means that there are 15 repeating process periods in the input series, and the series shows 15-day data, i.e., a specific process is repeated once per day.

### 3. Metrics anomaly detection

In our approach we proposed to use ML based algorithm that firstly learns how analyzed metrics behave in normal state and then is able to find anomaly behavior of time series. We tested 3 different algorithms with different parameters [4][5][6].

**Table 3.** Summary of tested variations of algorithms.

<b>Algorithm</b>	<b>Parameter</b>				
	<b>GPU (Computing power)</b>	<b>TIME STEPS</b>	<b>Loss (MAE vs MSE)</b>	<b>Dropout</b>	<b>Change in the number of neurons</b>
Autoencoder with 2 LSTM layers of 128 neurons each	Average run time for metrics is 37 s.	Min 50	Use MAE	Suggested 0.2-0.3	Not tested
Single layer LSTM network (128 neurons)	Runs 2x faster than alg. 1	Min 50	Use MAE	Not tested, unable to test	Min 32 neurons
Convolutional autoencoder	Runs 2x faster than alg. 1	Min 50	Use MAE	Suggested 0.2-0.3	Not tested



**Figure 3.** Blue line – value of the metric, the orange one – the loss function, the green one is a prediction, the red one is a threshold.

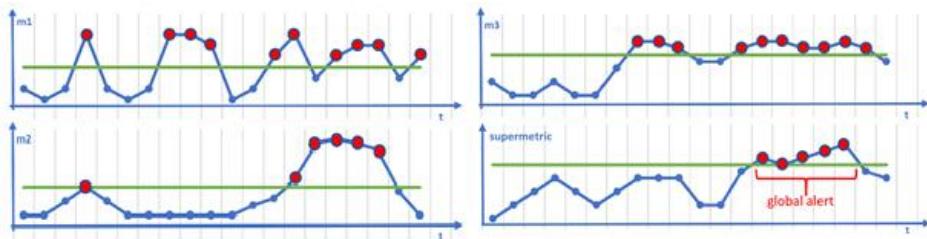
The purpose of the algorithm is to load data from metrics in order to detect anomalies. In the first cycle of the algorithm, the raw data of the metrics are loaded and divided in the ratio of 80/20 (training data and testing data). After the standardization and transformation of the data by the scaler, the function No. 1 (Autoencoder with 2 LSTM layers of 128 neurons each, with dropout 0.2 and TIME\_STEPS = 50) was used. This algorithm generates the most likely results after the consultation with the mainframe server administrators. Basing on these results the following approach was proposed: 1) calculate basing on the analysis of repeatability the value of TIME\_STEPS parameter, 2) basing on chosen ML algorithm data calculate the loss function as a difference between real data and predicted, and use the threshold as a max value of loss function. If the loss value is above the threshold generate the anomaly. The final results of this approach are seen in Fig. 3.

#### 4. Proposed solutions for event correlation detection in metrics

The so far proposed approach allowed for detection of anomalies in one metric, but there are many metrics and between them some correlations can exist. We can find them with several approaches. The simplest one is based on Pearson's linear correlation coefficient  $r$  [7], but the number of comparisons will be approximately  $n(n-1)/2$ , so for 700 time series, this is about 250,000 comparisons. This solution can produce a matrix that is symmetric about the main diagonal (values of 1 on the diagonal), which will indicate in the range -1 ... +1 the degree of correlation. A preliminary analysis took more than 16 hours to compute the matrix (available at [8]). The second approach can be based on the idea of supermetric - the metric of all metrics.

##### 4.1. ‘Supermetric’ - the metric of all metrics

As part of this approach, a solution is proposed to do a metric that synchronous records the current state of the system. If an anomaly occurred in several metrics ( $m_1, m_2, \dots, m_n$ ) the value of supermetric in the form of a mathematical formula will be calculated (see Fig. 4).



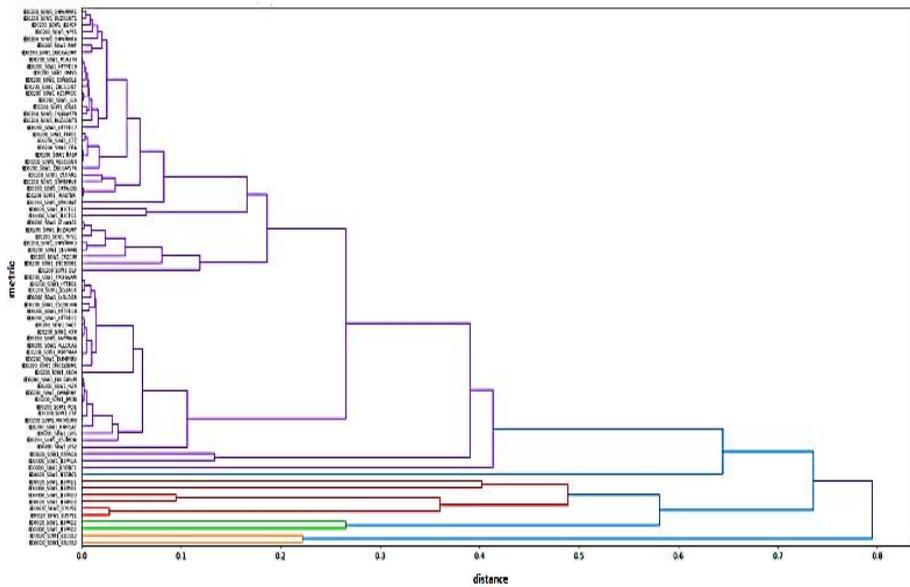
**Figure 4.** Overview drawing of the ‘supermetrics’. Green line is a threshold, red dots are anomalies.

Such a supermetric can be also analyzed via ML algorithm also to determine the trend of anomaly behavior (toward deepening anomalies, increasing their number, etc.). However, under this approach it is hard to find the exact mathematical formula for all existing metrics. In order to solve this problem we decided to use the following approach: having in mind that some metrics exhibit similar behavior we take the cluster analysis to find the clusters of metrics with similar statistical parameters (Fig. 5). According to [9] the key point is to find the right measures of distance that influences the final results. The carried experiments showed that the Ward method is the best one [10, 11].

Basing on these results we build for each cluster a new  $A_K(t)$  function (2) that for each cluster measures the possibility of second row anomaly in clusters:

$$A_K(t) = \frac{\sum_{n=1}^n \tanh(F_n)}{n}, \quad (2)$$

where:  $F_n$  – is the function of anomaly detection for each metric in the cluster,  $n$  – the number of metrics in the cluster,  $t$  – time.



**Figure 5.** The results of metrics clusterization.

For each cluster with defined  $A_K$  function we use the isolation forest algorithm to determine possible anomalies. The results are seen in Fig. 6.



**Figure 6.** Detection of second row anomalies in defined clusters.

## 5. Conclusions

In this paper we show some preliminary results of R&D activities related to the Z-RAYS support system for anomaly detection with application in mainframe management process. There is a lot of work to be done in the future comparing to the other similar solutions like Elastic Search [12] in order to guarantee that our proposed system will have the high specificity and sensitivity. The so far obtained results convince that proposed solutions can efficiently handle with the problem of automatic anomalies detection. The main research and development limitations are related to the computational power available for the whole support system. It is obvious that one can build a very sophisticated system that will require a huge amount of resources, however the results given by such a system will be no better than those guaranteed by a human operator.

## 6. Acknowledgements

This paper was partially supported by Polish National Centre for Research and Development (NCBiR) under agreement 9/2020, R&D project Z-RAYS – MAINFRAME SYSTEMS MONITOR.

## References

- [1] <https://www.ibm.com/docs/en/zos-basic-skills?topic=concepts-value-mainframe-today>
- [2] <https://www.ibm.com/products/z-development-test-environment>
- [3] H. Musbah, M. El-Hawary, H. Aly, Identifying Seasonality in Time Series by Applying Fast Fourier Transform, 2019 IEEE Electrical Power and Energy Conference (EPEC), 2019, pp. 1-4.
- [4] G. Hackeling. Mastering Machine Learning With scikit-learn. Packt Publishing. (2014).
- [5] F. T Liu, K. Ting, Z-H. Zhou, Isolation Forest. ICDM '08. Eighth IEEE International Conference on Data Mining, pp.413 - 422.(2009).
- [6] M. Togbe, et al. Anomaly Detection for Data Streams Based on Isolation Forest using Scikit-multiflow. The 20th Int. Conf. on Comp. Sci. and its Appl. (ICCSA 2020), Caligari, Italy.
- [7] J. Freeman and T. Young. Correlation coefficient: the association between two continuous variables. [https://www.sheffield.ac.uk/polopoly\\_fs/1.43991!/file/Tutorial-14-correlation.pdf](https://www.sheffield.ac.uk/polopoly_fs/1.43991!/file/Tutorial-14-correlation.pdf).
- [8] <https://github.com/pawkuras/SSFAD/blob/main/Table1.docx>
- [9] S. Saraklı, N. Doğan, I. Doğan, Comparison of hierarchical cluster analysis methods by cophenetic correlation. J. Inequal. Appl., 203 (2013).
- [10] S. Hands, B. Everitt. A Monte Carlo Study of the Recovery of Cluster Structure in Binary Data by Hierarchical Clustering Techniques. Multivariate Behav Res.22(2):235-43, (1987).
- [11] Murtagh, F., Legendre, P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?. J Classif 31, 274–295 (2014).
- [12] D. Turnbull, J. Berryman. Relevant search: with applications for Solr and Elasticsearch, Manning Publications Co., (2016).

# Ordering of Warships for the Brazilian Navy Using the New Method: AHP-Gaussian with Pearson's Correlation

Carlos Francisco Simões GOMES<sup>a,1</sup>, Marcus Vinícius Gonçalves RODRIGUES<sup>a</sup>, Igor Pinheiro de Araújo COSTA<sup>a</sup> and Marcos dos SANTOS<sup>b</sup>

<sup>a</sup>*Federal Fluminense University – UFF, Niterói, RJ, Brazil*

<sup>b</sup>*Military Engineering Institute – IME, Rio de Janeiro, RJ, Brazil.*

**Abstract.** This paper aims to support the selection decision of a medium-sized warship (between 2,000 and 3,000 tons), to be built in Brazil, presenting the alternatives in a hierarchical manner. Among the various multicriteria decision analysis (MCDA) methods, we used the analytic hierarchy process (AHP) as a basis. Throughout the study, we will propose some adjustments to the AHP in order to make the decision more robust (such as the use of the Gaussian factor and Pearson's correlation). The criteria were listed and their respective weights were assigned in light of the National Defense Strategy, the Navy's Strategic Program and interviews carried out with Brazilian navy officers with more than twenty years of career. To list the criteria, we adopted the critical incident technique. The use of the adapted AHP method in choosing the unit to be built can be considered as a transparent way, with a clearly scientific bias, for the Brazilian society to have the perception that the best option was made among the three models of warships presented.

**Keywords.** Multicriteria Decision Analysis; Analytic Hierarchy Process (AHP); Warship; Brazil's navy.

## 1. Introduction

According to the Brazilian national defense strategy [1], the science, technology and innovation policy for national defense aims to stimulate innovation, scientific and technological development of interest to the country's strategy. This happens through a national planning for the development of products with high technological content; with coordinated involvement of civil and military scientific and technological institutions, industry and universities; with the definition of priority areas and their respective technologies of interest; and the creation of instruments to encourage research into materials, equipment and systems for defense or dual employment. It is worth noting that all these actions seek to enable a technological and operational vanguard based on strategic mobility, flexibility and the ability to deter or surprise. Within this scope, despite other political and/or financial difficulties, the Brazilian navy (BN) has been trying to develop its own naval resources with its own technology.

---

<sup>1</sup> Corresponding Author, Carlos Francisco Simões GOMES, Federal Fluminense University – UFF, Niterói, RJ, Brazil; E-mail: cfsg1@bol.com.br

Brazil has a modern armed force equipped and prepared to guarantee its sovereignty and strategic interests, supporting its foreign policy and positions in international forums, even in peace times. The BN is supposed to have 18 warships (including frigates, corvettes and destroyers) in order to permit the arrangement of two task groups (one with close escort to a main body and another for the away defense of the surface action group). Escort ships must have anti-submarine, anti-surface and anti-air point defense capabilities (missile availability). They must have the capacity to transport, supply, operate and maintain helicopters, capable of attacking surface and submarine targets, in addition to carrying out clarification operations.

Since the last decade of the 20th century, the budget made available by the Brazilian Federal Government has fallen short of the amount needed to meet the needs of the BN, making it almost impossible to allocate the necessary and sufficient credits for operation, maintenance, and re-equipment. The armed forces, especially the BN, have suffered budgetary restrictions that is resulting in not meeting the minimum needs of the naval force, also resulting in operational capacity reduction and in the accumulation of demands. According to Almeida et al. [2], Moreira et al. [3] and Tenorio et al. [4], due to progressive budget restrictions, nowadays, the BN has an undersized fleet of only 11 escort ships (many of which are close to the conclusion of their useful live), resulting in non-compliance with the minimum necessities of the naval force, diminishing the operational capacity and with the accumulation of requests from different sources.

Gomes et al. [5], Oliveira et al. [6] and Rodrigues et al. [7], highlight in their work that, in general, selection problems must be modeled by multicriteria methods. In this context, this work aims to select a new warship to be built by the BN through the application of the Analytic Hierarchy Process (AHP), which is one of the most classic multiple criteria decision support method (from the American school). The AHP was chosen because it is a compensatory and hierarchical method mainly indicated for problems with an average number of alternatives and criteria, considering the discrimination of results and cognitive effort in paired comparisons. According to Costa et al. [8], Jardim et al. [9] and Moreira et al. [10], the concepts of hierarchy and compensatory decision rules are in accordance with the military culture, which facilitates the analysis by specialists. We also proposed a new approach to the AHP method in order to make the decision more robust (such as the use of the Gaussian factor and Pearson's correlation).

## 2. Literature Review

AHP, developed by Saaty [11], is one of the most classic methods of the American school (non-compensatory methods) to quantitatively treat complex problems with multiple decision criteria. The essence of the method lies in the weighting of the criteria weights, by comparing two by two, using a metric known as: Saaty's fundamental scale, which can be found in Table 1.

**Table 1.** Saaty's Fundamental Scale

Reference	Meaning	Explanation
1	Equal importance	The two alternatives contribute equally to the goal.
3	Moderate importance	Experience and judgment slightly favor one activity over another.
5	Strong importance	Experience and judgment strongly favor one activity over another.

7	Very strong importance	One alternative is strongly favored over another.
9	Extreme importance	The evidence is in favor of one alternative over another (to the greatest extent possible)
2,4,6,8	Intermediate values	Used to express preferences that are between the references above.

Source: Adapted from Saaty [4]

According to Saaty [11], in order to ensure that the decision has consistency, the AHP method calculates the Consistency Ratio (CR) between the Consistency Index (CI) of the judgments and the Random Consistency Index (RI). The maximum tolerance for inconsistency allowed by the model is ten percent, so  $CR \leq 0.1$ . If  $CR > 0.1$ , the decision maker must make a new evaluation of the criteria. Below is a summary of the AHP method according to Santos et al. [12] and Souza et al. [13]:

a) Formation of the decision matrices ( $m \times n$ ) (1):

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \quad (1)$$

The decision matrices express the number of times an alternative dominates or is dominated by the others.

b) Calculation of the eigenvector ( $W_i$ ) (2):

$$W_i = \left( \prod_{j=1}^n W_{ij} \right)^{1/n} \quad (2)$$

This step consists in ordering the priorities or hierarchies of the studied characteristics.

c) Calculation of eigenvectors normalization, enabling comparability between criteria and alternatives (3):

$$T = \left| \frac{w_1}{\sum w_i} ; \frac{w_2}{\sum w_i} ; \frac{w_3}{\sum w_i} ; \dots ; \frac{w_n}{\sum w_i} \right| \quad (3)$$

d) Determination of the index that relates the criteria of the consistency matrix and the criteria weights (4):

$$\lambda_{max} = T \times W \quad (4)$$

e) Determination of the consistency index (CI) (5):

$$CI = \frac{\lambda_{max} - n}{(n - 1)} \quad (5)$$

This step allows the decision maker to assess the degree of inconsistency of the paired judgments matrix.

f) Calculation of the consistency ratio (CR) (6):

$$CR = \frac{CI}{RI} \quad (6)$$

Where, RI is the random consistency index obtained for a reciprocal matrix of order  $n$ . This step allows evaluating the inconsistency due to the order of the judgment matrix. If the CR value is greater than 0.10, the decision maker must review the model or judgments.

### 3. Methodology

The development of this research took place, in summary, through the steps below:

- i.Identification and definition of the problem with the BN;
- ii.After defining the problem, we performed a literature review to contextualize the study and map the main methods used by the authors;
- iii.Definition and analysis of the main technical characteristics of the three warships that could be built by the BN;
- iv.Definition of the mathematical modeling tool to be used (in this case, we choose the AHP);
- v.Interview with ten BN officers (with more than twenty years of operational experience) in order to list the relevant criteria for choosing the most suitable warship for the Brazilian navy's needs, as well as establishing the inter-criteria weights;
- vi.Structuring the problem in the mental map form based on the concept established by specialists who are knowledgeable about the warship process of operation and maintenance;
- vii.Definition of the final criteria for choosing the warship;
- viii.Modeling the problem in a spreadsheet;
- ix.After ranking by AHP, we identified a possibility of improving the method (bringing more robustness to the choice of the first alternative). The improvement consisted in the correction of the weights modulus by using the Gaussian factor and the Pearson's correlation;
- x.Ranking of alternatives and comparison of results with results obtained in traditional AHP.

### 4. Data Analysis and Results

The BN started in 1994 the construction of the Corvette Barroso and finished the construction in 2008. Corvette Barroso is a medium-sized 2,500-ton warship that was launched only 14 years after the start of work (thus, they had a new ship in their arsenal but not modern compared to other models available on the market).

Nowadays, with the need to build new warships to compose the fleet, the BN is facing a new challenge, which is: to replicate the model of the current Corvette Barroso, build a slightly modernized ship (2,600 ton corvette) or build a model with more significant upgrades (3,000 ton corvette). It is important to highlight that the decision-making process must meet the requirements of the BN, also prioritizing cost reduction. We collected the three warships' data, with the main technical characteristics of the ships, from Vogt's [14] study.

After the interview with experts from the BN, nine criteria were considered for choosing the warship to be built. They are listed below:

(C<sub>1</sub>) Action radius: greatest distance (in nautical miles) the ship can travel from its base and return without refueling;

(C<sub>2</sub>) Fuel endurance: time interval (in days) that a ship can navigate without refueling with speed at 15 knots.

(C<sub>3</sub>) Autonomy: maximum interval of time (in days) that a ship can operate without any type of supplies (fuel, drinking water, foodstuff, etc.).

(C<sub>4</sub>) Primary cannon: a weapon with a high rate of fire that functions to warn or neutralize possible threats, such as ships, aircraft or missiles. It is called "primary" when the ship has other alternative guns, usually of smaller caliber.

(C<sub>5</sub>) Secondary cannon: an alternative cannon to the "primary cannon", usually of smaller caliber.

(C<sub>6</sub>) AAW missiles: anti-aircraft warfare missiles.

(C<sub>7</sub>) Initial Cost: cost of obtaining or building a ship.

(C<sub>8</sub>) Life Cycle Cost: life cycle cost of a ship, includes the purchase (or construction), operation and modernization. The purchase price represents about 25%, the expenses for crew and operations account for 67%, and the possible modernization corresponds to 8%.

(C<sub>9</sub>) Construction Time: this criterion is self-explanatory, considering from the start of the project to the actual delivery of the ship to the operating sector.

At first, we applied the traditional AHP method to our problem. Table 2 illustrates the problem decision matrix and its normalization.

**Table 2.** Decision Matrix and its normalization

Type	Reference	(1)	(2)	(3)
Decision Matrix	C1	4000	9330	10660
	C2	11	26	30
	C3	30	25	35
	C4	25	25	120
	C5	1	2	2
	C6	0	1	1
Normalized Decision Matrix	C7	R\$ 290,000,000	R\$ 310,000,000	R\$ 310,000,000
	C8	R\$ 592,000,000	R\$ 633,000,000	R\$ 633,000,000
	C9	6 years	8 years	8 years
	C1	0.1667	0.3889	0.4443
	C2	0.1641	0.388	0.4477
	C3	0.3333	0.2777	0.3888
	C4	0.147	0.147	0.7058
	C5	0.2	0.4	0.4
	C6	0	0.5	0.5
	C7	0.3483	0.3258	0.3258
	C8	0.3483	0.3258	0.3258
	C9	0.4	0.3	0.3

Source: Done by the authors.

Table 3 illustrates the weight matrix and its normalization.

**Table 3.** Weight Matrix and its Normalization

Type	Reference	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>	C <sub>5</sub>	C <sub>6</sub>	C <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>
Weight Matrix	C <sub>1</sub>	1	1	1	0.33	5	1	0.33	0.33	0.25
	C <sub>2</sub>	1	1	1	0.33	5	1	0.33	0.33	0.25
	C <sub>3</sub>	1	1	1	0.33	5	1	0.33	0.33	0.25
	C <sub>4</sub>	3	3	3	1	3	1	0.33	0.33	0.33
	C <sub>5</sub>	0.20	0.2	0.2	0.33	1	0.2	0.2	0.2	0.14
	C <sub>6</sub>	1	1	1	1	5	1	0.33	0.33	0.33

	C <sub>7</sub>	3	3	3	3	5	3	1	1	1
	C <sub>8</sub>	3	3	3	3	5	3	1	1	1
	C <sub>9</sub>	4	4	4	3	7	3	1	1	1
	C <sub>1</sub>	0.06	0.06	0.06	0.03	0.12	0.07	0.07	0.07	0.05
	C <sub>2</sub>	0.06	0.06	0.06	0.03	0.12	0.07	0.07	0.07	0.05
	C <sub>3</sub>	0.06	0.06	0.06	0.03	0.12	0.07	0.07	0.07	0.05
Normalized Weight Matrix	C <sub>4</sub>	0.17	0.17	0.17	0.08	0.07	0.07	0.07	0.07	0.07
	C <sub>5</sub>	0.01	0.01	0.01	0.03	0.02	0.01	0.04	0.04	0.03
	C <sub>6</sub>	0.06	0.06	0.06	0.08	0.12	0.07	0.07	0.07	0.07
	C <sub>7</sub>	0.17	0.17	0.17	0.24	0.12	0.21	0.20	0.20	0.22
	C <sub>8</sub>	0.17	0.17	0.17	0.24	0.12	0.21	0.20	0.20	0.22
	C <sub>9</sub>	0.23	0.23	0.23	0.24	0.17	0.21	0.20	0.20	0.22

Source: Done by the authors.

Table 4 illustrates the results found after applying the traditional AHP method.

**Table 4.** Results (traditional AHP)

Classification	Alternative	Score
1 <sup>st</sup> option	(3) totally new model	0.3949
2 <sup>nd</sup> option	(2)	0.3207
3 <sup>rd</sup> option	(1) current model	0.2843

Source: Done by the authors.

Analyzing the results shown in Table 4, after applying the traditional AHP method, we can see that the best construction alternative, even considering the risks involved, is model (3). Meanwhile, the worst alternative is replicating the current model (1). The traditional AHP method generated a small discrimination among alternatives, which may indicate the need for a more careful sensitivity analysis. To increase the discriminatory power, we propose the following: to calculate the standard deviation of the alternatives for each criterion, to calculate de Gaussian factor for each criterion, to calculate the Pearson's correlation module among the criteria in order to have a greater discrimination of the criteria weights and compensate the weights generated by the AHP-Gaussian through the values generated by the Pearson correlation. The new results can be found in Table 5.

**Table 5.** Results (New approach to the AHP – Gaussian Factor + Pearson's Correlation)

Classification	Alternative	Score
1 <sup>st</sup> option	(3) totally new model	0.5193
2 <sup>nd</sup> option	(2)	0.2750
3 <sup>rd</sup> option	(1) current model	0.2057

Source: Done by the authors.

It is possible to notice that we obtained the same ranking in both methods. By using the Gaussian factor and the Person's correlation to greater differentiate the criteria weights, we were able to further legitimize that model 3 is the best option to be built.

## 5. Conclusion

For over one hundred and thirty years, Brazil has lived at peace with its neighbors, however, this does not mean the end of external threats. History clearly demonstrates the need for a nation to have naval power that inspires credibility. In general, in most

countries it is up to the naval power to guarantee the integrity of the territory, national sovereignty and defend the interests of the nation in its waters, aiming to ensure the right to the strategic and economic use of the sea and inland waters. Due to the phenomenon of globalization, the struggle to defend national interests, whether political, economic or strategic military, paradoxically, has been much more forceful than in past decades. Opposing globalization is not productive, as it tends to expand more and more, with implications for the performance of the armed forces. There is a time when many solutions from the past no longer apply. There is a need to evolve in organizational terms and management methods.

Thus, this article proposes a new approach to the AHP method by combining the Gaussian factor and the Pearson's correlation in order to have a greater differentiation among the alternatives. Through interviews carried out with specialists from the BN, we elicited the weights and ranked the potential warships to be built by the BN. The conclusion was that, for the Brazilian scenario, the best solution is to build a completely new ship, which was later corroborated through a sensitivity analysis. For future studies, the methodology used in this article can be adapted to other contexts: in other industries, in other countries, with different databases. It is also interesting to apply different methods to prove the robustness of the results found, as we did in this article. The results found in this paper were forwarded to the Brazilian navy.

## References

- [1] Ministério da Defesa do Brasil. Estratégia Nacional de Defesa. Brasília: Ministério da Defesa do Brasil, 2016. 11 pages. Available in: [https://www.gov.br/defesa/pt-br/arquivos/estado\\_e\\_defesa/pnd-end-2016.pdf](https://www.gov.br/defesa/pt-br/arquivos/estado_e_defesa/pnd-end-2016.pdf). Accessed in: 08 fev. 2021.
- [2] ALMEIDA, Isaque David Pereira de; CORRÍÇA, José Victor de Pina; COSTA, Arthur Pinheiro de Araújo; COSTA, Igor Pinheiro de Araújo; MAEDA, Sérgio Mitihiro do Nascimento; GOMES, Carlos Francisco Simões; SANTOS, Marcos dos. Study of the Location of a Second Fleet for the Brazilian Navy: structuring and mathematical modeling using sapevo-m and vikor methods. *Production Research*, [S.L.], p. 113-124, 2021. Springer International Publishing. [http://dx.doi.org/10.1007/978-3-030-76310-7\\_9](http://dx.doi.org/10.1007/978-3-030-76310-7_9).
- [3] MOREIRA, Miguel Ângelo Lellis; COSTA, Igor Pinheiro de Araújo; PEREIRA, Maria Teresa; SANTOS, Marcos dos; GOMES, Carlos Francisco Simões; MURADAS, Fernando Martins. PROMETHEE-SAPEVO-M1 a Hybrid Approach Based on Ordinal and Cardinal Inputs: multi-criteria evaluation of helicopters to support brazilian navy operations. *Algorithms*, [S.L.], v. 14, n. 5, p. 140, 27 abr. 2021. MDPI AG. <http://dx.doi.org/10.3390/a14050140>
- [4] TENÓRIO, Fabricio Maione; SANTOS, Marcos dos; GOMES, Carlos Francisco Simões; ARAUJO, Jean de Carvalho. Navy Warship Selection and Multicriteria Analysis: the thor method supporting decision making. *Industrial Engineering And Operations Management*, [S.L.], p. 27-39, 2020. Springer International Publishing. [http://dx.doi.org/10.1007/978-3-030-56920-4\\_3](http://dx.doi.org/10.1007/978-3-030-56920-4_3)
- [5] GOMES, Carlos Francisco Simões; SANTOS, Marcos dos; TEIXEIRA, Luiz Frederico Horácio de Souza de Barros; SANSEVERINO, Adriana Manzolillo; BARCELOS, Mara Regina dos Santos. SAPEVO-M: a group multicriteria ordinal ranking method. *Pesquisa Operacional*, [S.L.], v. 40, p. 1-23, 2020. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/0101-7438.2020.040.00226524>
- [6] OLIVEIRA, Altina S.; GOMES, Carlos F. S.; CLARKSON, Camilla T.; SANSEVERINO, Adriana M.; BARCELOS, Mara R. S.; COSTA, Igor P. A.; SANTOS, Marcos. Multiple Criteria Decision Making and Prospective Scenarios Model for Selection of Companies to Be Incubated. *Algorithms*, [S.L.], v. 14, n. 4, p. 111, 30 mar. 2021. MDPI AG. <http://dx.doi.org/10.3390/a14040111>
- [7] RODRIGUES, Marcus Vinícius Gonçalves; DUARTE, Thamires Eis; SANTOS, Marcos dos; GOMES, Carlos Francisco Simões. Prospective scenarios analysis: impact on demand for oil and its derivatives after the covid-19 pandemic. *Brazilian Journal Of Operations & Production Management*, [S.L.], v. 18, n. 2, p. 1-15, jun. 2021. Associação Brasileira de Engenharia de Produção - ABEPROM. <http://dx.doi.org/10.14488/bjopm.2021.039>. Available in: <https://bjopm.emnuvens.com.br/bjopm/article/view/1215/998>. Accessed on 03 apr. 2021.

- [8] COSTA, Igor Pinheiro de Araújo; MAEDA, Sérgio Mitihiro do Nascimento; TEIXEIRA, Luiz Frederico Horácio de Souza de Barros; GOMES, Carlos Francisco Simões; SANTOS, Marcos dos. Choosing a hospital assistance ship to fight the covid-19 pandemic. *Revista de Saúde Pública*, [S.L.], v. 54, p. 79, 10 ago. 2020. Universidade de São Paulo, Agencia USP de Gestao da Informacao Academica (AGUIA). <http://dx.doi.org/10.11606/s1518-8787.2020054002792>
- [9] JARDIM, R. R. J.; SANTOS, M.; E. NETO,; SILVA, E. da; BARROS, F. de. Integration of the waterfall model with ISO/IEC/IEEE 29148: 2018 for the development of military defense system. *Ieee Latin America Transactions*, [S.L.], v. 18, n. 12, p. 2096-2103, dez. 2020. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/tla.2020.9400437>
- [10] MOREIRA, Miguel Ângelo Lellis; GOMES, Carlos Francisco Simões; SANTOS, Marcos dos; SILVA, Marcela do Carmo; ARAUJO, Jonathas Vinícius Gonzaga Alves. PROMETHEE-SAPEVO-M1 a Hybrid Modeling Proposal: multicriteria evaluation of drones for use in naval warfare. *Industrial Engineering And Operations Management*, [S.L.], p. 381-393, 2020. Springer International Publishing. [http://dx.doi.org/10.1007/978-3-030-56920-4\\_31](http://dx.doi.org/10.1007/978-3-030-56920-4_31).
- [11] Saaty, T. L. (1980) *Analytic Hierarchy Process*. New York: McGraw Hill.
- [12] SANTOS, M.; COSTA, I. P. ARAUJO; GOMES, C. F. S (2021). Multicriteria Decision-Making In The Selection Of Warships: A New Approach To The AHP Method. *International Journal of the Analytic Hierarchy Process*, v. 13, p. 1-13.
- [13] DE SOUZA, L. P.; GOMES, C. F. S.; DE BARROS, A. P. (2018). Implementation of new Hybrid AHP-TOPSIS-2N method in sorting and prioritizing of an it capex project portfolio. *International journal of information technology & decision making*, v. 17, p. 977-1005.
- [14] Vogt, R. (2013). Corvetas Sucessoras da Barroso: comparação de dois tipos de obtenção. *Revista Marítima Brasileira*, Rio de Janeiro, 133(04/06), 108–129.

# Strategic Analysis for the Installation of Field Hospitals for COVID-19 Control: An Approach Based on P-Median Model

Leandro de Mattos Bento Soares<sup>a</sup>, Miguel Ângelo Lellis Moreira<sup>a,b,1</sup>,  
 Marcio Pereira Basilio<sup>c</sup>, Carlos Francisco Simões Gomes<sup>b</sup>,  
 Marcos dos Santos<sup>a,d</sup>, and Igor Pinheiro de Araújo Costa<sup>a,b</sup>

<sup>a</sup>*Naval Systems Analysis Center, Rio de Janeiro, RJ 20091-000, Brazil*

<sup>b</sup>*Fluminense Federal University, Niterói, RJ 24210-240, Brazil*

<sup>c</sup>*Military Police of the State of Rio de Janeiro, Rio de Janeiro, 20031-040, Brazil*

<sup>d</sup>*Military Institute of Engineering, Urca, RJ 22290-270, Brazil*

**Abstract.** Concerning the expansion of the coronavirus in the world, the search for the development of solutions related to the control and mitigation of the pandemic situation became constant. The paper addresses an analysis of localities for the installation of field hospitals, highly requested as a point of treatment for COVID-19. In this scenario, a framework based on the P-median approach and mathematical programming is proposed, enabling an optimization model as an analysis format for the problematic situation. To support the implementation of the model, a computational tool for data processing was developed, integrating an optimization model to the geographical evaluation, exploring in the analysis numerical and graphical resources. As a validation of the study, a case study in the city of Rio de Janeiro – Brazil is presented, analyzing 162 neighborhoods and determining seven favorable localities for the installation of field hospitals.

**Keywords.** P-median, Mathematical Programming, Computational Tool.

## 1. Introduction

The COVID-19 global pandemic was easily spread by the proximity of people, especially in large agglomerations, a common characteristic of big cities and metropolitan regions [1]. Since its dissemination, the coronavirus has highlighted a series of weaknesses in the health systems around the world, highlighting the problems in accessing the essential resources needed to face this pandemic, such as the lack of respirators and protective equipment [2].

Regarding the expansion of the disease, facilitated by its transmission through contact and movement of people, it was necessary to adopt isolation measures that help in the control and mitigation of virus transmission [3]. In this context, the search for alternative solutions motivated the development of studies in the most diverse areas of science, in search of providing knowledge that would make prevention, prediction, and consequently the minimization and control of the COVID-19 pandemic [4].

---

<sup>1</sup> Corresponding Author, Miguel Ângelo Lellis Moreira, Production Department, Fluminense Federal University, Niterói, RJ 24210-240, Rio de Janeiro, Brazil; E-mail: miguellellis@id.uff.br.

In this scenario, in many cities, there was a need to create field hospitals based on temporary facilities to carry out diagnostic tests, hospitalization, and initiation of early treatments for the population, searching to mitigate the problems related to the given pandemic moment [5]. Based on this fact, was made the following question: How to define a favorable location for the installation of field hospitals, minimizing circulation of the population, and providing access to medical care?

Concerning the questioning, the Operational Research (OR) provides, through its approaches and methodologies, the analysis of complex problems with the technical and scientific basis, providing the structuring, understanding, and, in specific cases, the viability of an optimal or favorable solution for a given problem [6].

The problem of optimal allocation of resources is one of the main areas of study present in OR, thus mathematical programming models, aimed at optimizing resources, minimizing or maximizing, make it possible to adapt to the problem in question [7]. In this context, the given models can work as an aid in the clarification of favorable points for the installation of campaign hospitals based on a set of restrictions present in the context and location under analysis.

Regarding the scenario presented, the study addresses the evaluation of localities in the city of Rio de Janeiro – Brazil for the installation of field hospitals as an aid in the control of the COVID-19 pandemic. For the analysis, an optimization model based on P-median modeling is provided, enabling the identification of favorable localities to install the field hospitals. In search to support the implementation, a computational framework based on R software [8] was developed, making possible the implementation of the optimal algorithm, by the moment that dealing with the analysis of big data sets, computational tools make possible not only to explore numerical resources but also graphical resources in favor of a more interactive analysis with the decision-maker [9].

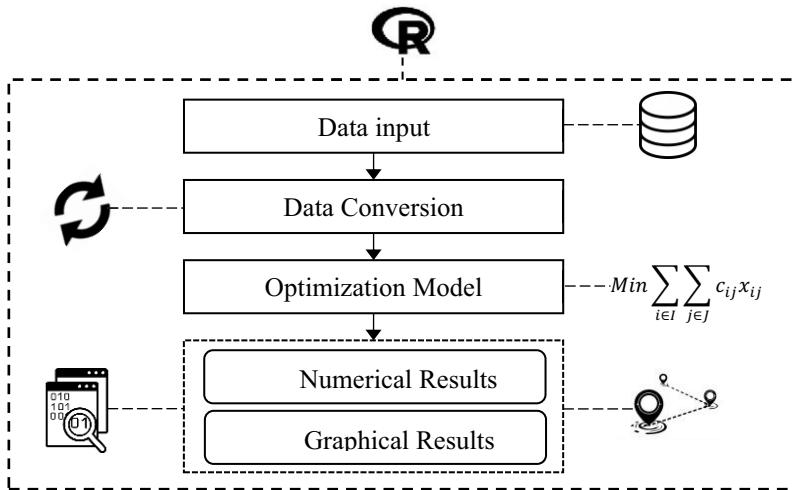
The paper is structured into four sections. After the introduction, section 2 presents the developed framework, exploring the mathematical modeling and computational resources. Section 3 presents the case study, detailing the problematic situation and data set and exploring the steps of the numerical procedure and the results analysis. Section 4 concludes the study, exposing the gains of the approaching and proposal for future works.

## 2. Framework Model Analysis

As discussed in the introduction, the framework presents the integration between mathematical programming, data processing, and graphical analysis. The computational tool was developed in R language [8], which stands out among technologies based on mathematical and statistical applications, enabling to perform numerical and graphical analysis through your packages and tools [10].

The proposed approach allows for flexibility regarding the restrictions inputted in the mathematical programming model. However, it is worth mentioning that, in general, it uses modeling to solve the P-median problem. It is noteworthy that the framework, is restricted to the geographical dataset, that is, the input data consists of latitudinal and longitudinal points.

The script provides a simple and intuitive interaction, firstly, the structuring and conversion of the dataset based on latitudinal and longitudinal data is performed. Then, the mathematical modeling of the problem to implement the optimization model is realized. In the end, the results through numerical tables and geographical graphics are exposed. Figure 1 presents the flowchart concerning the framework procedure.

**Figure 1.** Framework optimal model

## 2.1. Data Processing

The given step consists of the conversion of the dataset to numerical rating scores to set up the optimization model. Thus, initially, a dataset (.csv) based on the information presented in table 1 is imported, exposing all possible installation points, describing their location and geographic coordinates. Considering that the framework is related to the analysis of the COVID-19 pandemic situation, the number of infected cases of each region works as a weighting factor for the problem.

**Table 1.** Data set input description

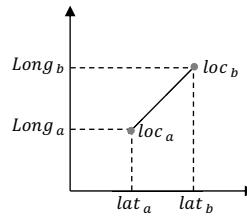
Facilities	Localities	Latitude	Longitude	Covid Cases
Field Hospital <sub>1</sub>	Place <sub>1</sub>	Lat <sub>1</sub>	Long <sub>1</sub>	$nc_1$
Field Hospital <sub>2</sub>	Place <sub>2</sub>	Lat <sub>2</sub>	Long <sub>2</sub>	$nc_2$
Field Hospital <sub>n</sub>	Place <sub>n</sub>	Lat <sub>n</sub>	Long <sub>n</sub>	$nc_n$

Localization problems can arise in two different scopes: the plane and the network, considering the necessity to express the distance between a pair of points. The network handles all possible paths between each pair of points, which may eventually be affected by topographical elements, barriers, and oriented paths. On the other hand, the plane uses the concept of Euclidean distances, which defines the shortest distance between two points based on a straight line connecting them [11].

Thus, as presented in [12], to define the Euclidean distance of  $(A, B)$ , given by the points  $A = (x_a, y_a)$  and  $B = (x_b, y_b)$ , considering that the segment formed by points  $A$  and  $B$  is not parallel to any of the axes, it is used a point  $C = (x_b, y_a)$ , and points  $A, B$  and  $C$  become vertices of a right triangle in  $C$ , applying the Pythagorean Theorem (1).

$$d_{ab}(A, B) = \sqrt{(x_b - x_a)^2 + (y_b - y_a)^2} \quad (1)$$

Based in the given problematic, the equation (1) is operated by latitudinal and longitudinal points, as presented in figure 2. The number of covid cases weights the function, generating a rating scores ( $c_{ab}$ ) (2) for each variable of optimization problem.



**Figure 2.** Distance between two localities

$$c_{ab} = d_{ab}(A, B) \cdot nc_b \quad (2)$$

Obtained the rating scores ( $c_{ab}$ ) between the pairs of locations, a matrix of variables (3) is generated, exposing the respective costs among all locations of the problem.

$$\begin{pmatrix} c_{11} & c_{12} & c_{13} & \dots & c_{1n} \\ c_{21} & c_{22} & c_{23} & \dots & c_{2n} \\ c_{31} & c_{32} & c_{33} & \dots & c_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & c_{n3} & \dots & c_{nn} \end{pmatrix} \quad (3)$$

## 2.2. Optimization Model

Based on the *P-median* model [13], the modeling consists of a mathematical model of integer programming, in a binary format, searching for determining optimal points within a set of constraints [14,15], as exposed by system (4).

$$\begin{aligned} \text{Min } & \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} \\ \text{s.t. } & \sum_{i \in I} x_{ij} = 1 \quad \forall j \in J \\ & x_{ij} \leq y_i, \quad \forall i \in I, \quad \forall j \in J \\ & \sum_{i \in I} y_i = P \\ & x \in B^{|I||J|}, \quad y \in B^{|I|} \end{aligned} \quad (4)$$

In the given system, the objective function minimizes the total rating scores of the evaluation, indicating the minimum possible agglomeration of the population to access a field hospital. The restrictions certify that each location will be served by only one location, and can serve itself. The variable  $y_i$  presents the number of points available for service, limited to  $P$ , which indicates the maximum possible installations.

Concerning the computational tool, the Lpsolve package [16] was implemented, making possible to perform optimal algorithms based on mathematical programming.

## 2.3. Model of Result Analysis

In search to provide a robustness analysis, integrated with numerical resources, it is enabled on the framework a geographical analysis of problematic. Through the implementation of Leaflet package [17], it is possible to plot the points in evaluation in a real map, built by geographic data.

After provide the optimal results, the favorable points for facilities are exposed in the map, highlighting the facility centers and their respective locations for access, providing an interactive graphical analysis of results.

### 3. Case Study

Using the city of Rio de Janeiro as a region of analysis for the given problem a set of neighborhoods that provide the installation of a limited number of campaign hospitals in the fight against COVID-19 is considered. All data presented in the study were provided through data repositories of the Brazilian Ministry of Health, and are available for public access (<https://covid.saude.gov.br/>).

In the evaluation, 162 locations were considered, where their respective geographic coordinates using the Google Earth Pro software were obtained. For study purposes, the data regarding the numbers of infected in each region in a constant format is performed. Table 2 presents partially the set of inputs.

**Table 2.** Partial set of input

Facilities	Localities	Latitude	Longitude	Covid Cases
Field Hospital <sub>1</sub>	Abolição	-22.888002	-43.299279	289
Field Hospital <sub>2</sub>	Acari	-22.825531	-43.344460	177
Field Hospital <sub>3</sub>	Água Santa	-22.905568	-43.308167	180
:	:	:	:	:
Field Hospital <sub>162</sub>	Vista Alegre	-22.8303	-43.3167	365

Based on the set of inputs, the algorithm developed (Figure 1) establishes the variable costs, representing the weighted value of Euclidean distances by the number of infected cases per region, as presented in section 2.1. Table 3 presents the partial matrix of costs generated for each pair of locations.

**Table 3.** Partial set of costs between each pair of localities

Localities	Abolição	Acari	Água Santa	Anchieta	...	Vista Alegre
Abolição	0	13.6	3.50	102.7	...	5.60
Acari	22.3	0	15.8	45.60	...	6.70
Água Santa	5.70	15.6	0	106.4	...	6.30
Anchieta	34.2	9.30	22.1	0	...	8.80
:	:	:	:	:	..	:
Vista Alegre	40.4	29.7	28.1	41.3	...	0

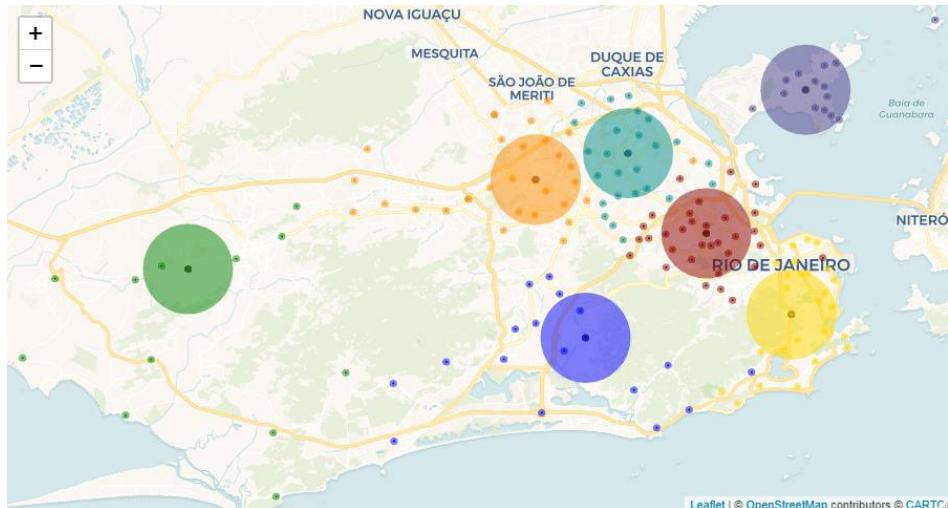
In search to determine the optimal locations for the installation of field hospitals, the optimization model was applied. Through the framework, 26406 variables ( $x_{ij}$ ) corresponding to 162 localities ( $y_i$ ) were generated. In the evaluated scenario, a maximum of seven field hospitals ( $P = 7$ ) were considered.

After the processing of the optimal model, the seven favorable localities to the installation of field hospitals at Rio de Janeiro were defined. Table 4 exposes the optimal localities, along with the number and percentage of access localities.

**Table 4.** Partial set of input

Facilities	Localities	locations served (nº)	locations served (%)
Field Hospital <sub>1</sub>	Anil	18	11.1 %
Field Hospital <sub>39</sub>	Cosme Velho	26	16 %
Field Hospital <sub>68</sub>	Inhoaíba	13	8 %
Field Hospital <sub>72</sub>	Jacaré	33	20.37 %
Field Hospital <sub>77</sub>	Jardim Carioca	15	9.26 %
Field Hospital <sub>93</sub>	Marechal Hermes	31	19.13 %
Field Hospital <sub>155</sub>	Vila da Penha	26	16 %

For obtained results, it was possible to generate the graphical analysis by geographical map. As presented in figure 3, it was defined the seven optimal locations, each one represented for one color. The center point of each colored circle represents the optimal locality, and the colored points marked on the map represent the other localities of the evaluated set, that should access the optimal location for their respective color.



**Figure 3.** Graphical results of the optimal localities

Considering the rating scores to be minimized ( $c_{ab}$ ), representing a relative agglomeration of people to access field hospitals, in the evaluation a total value of 1663.1 was obtained. In addition to the analysis, three other analyses, now considering a maximum ( $P$ ) of 10, 15, and 20 field hospitals, resulting in rating scores of 1307, 923.5, and 719.7, respectively, were carried out. In this scenario, is understood the decreasing of rating scores by the increase of installation of field hospitals.

With the given analysis, it is observed that in addition to making it possible to determine optimal locations, it is also possible to clarify a favorable number of field hospitals, searching to find a balance point for a given total number of hospitals.

Based on the realized analysis, it is necessary to emphasize that the given model, in addition to providing a numerical, geographical, and graphical analysis based on the P-median problem, can be adapted to different case studies and data sets, with the need to obtain the coordinates geographic locations of a given problem.

#### 4. Conclusion

The study approaches the analysis of localities for the installation of field hospitals in the fight against COVID-19, using the P-median modeling. In a case study, the city of Rio de Janeiro was used as a basis for evaluation, identifying the possible neighborhoods to be used in the construction of facilities concerning a pandemic situation.

For a given analysis, the creation of rating scores variables indicating a relative agglomeration of the population to access a given medical care point was used. It is noteworthy that the given model brings an approximate analysis, with updates in future studies, however, it can be of great value for decision-making scenarios.

A computational tool to support the evaluation process was developed, enabling the conversion of geographical data to build an optimization model based on mathematical programming and the P-median model. Based on the results generated by the optimization model, numerical and graphical resources of result analysis were explored.

In this context, it concludes that the given study is favorable to the analysis of localities for facilities, and the computational framework can be adapted to different types of analyzes and problems. For future works, we search to integrate other constraining factors into the optimization model, making it possible to consider different service capabilities, and other resource limitations providing a closer approximation of the model to reality.

## References

- [1] Alsaeedy AAR, Chong EKP. Detecting Regions At Risk for Spreading COVID-19 Using Existing Cellular Wireless Network Functionalities. *IEEE Open Journal of Engineering in Medicine and Biology* 2020;1:187–9. <https://doi.org/10.1109/ojemb.2020.3002447>.
- [2] de Sutter A, Llor C, Maier M, Mallen C, Tatsioni A, van Weert H, et al. Family medicine in times of ‘COVID-19’: A generalists’ voice. *European Journal of General Practice* 2020;26:58–60. <https://doi.org/10.1080/13814788.2020.1757312>.
- [3] Costa IP de A, Sanseverino AM, Barcelos MR dos S, Belderrain MCN, Gomes CFS, Santos M dos. Choosing flying hospitals in the fight against the COVID-19 pandemic: structuring and modeling a complex problem using the VFT and ELECTRE-MOr methods. *IEEE Latin America Transactions* 2021;19:1099–106. <https://doi.org/10.1109/TLA.2021.9451257>.
- [4] Costa IPA, Maeda SMDN, Teixeira LFHSB, Gomes CFS, Santos MD. Choosing a hospital assistance ship to fight the covid-19 pandemic. *Revista de Saude Publica* 2020;54:79. <https://doi.org/10.11606/s1518-8787.2020054002792>.
- [5] Rodrigues MVG, Duarte TE, dos Santos M, Gomes CFS. Prospective scenarios analysis: Impact on demand for oil and its derivatives after the COVID-19 pandemic. *Brazilian Journal of Operations & Production Management* 2021;18:e20211215. <https://doi.org/10.14488/BJOPM.2021.039>.
- [6] Moreira MÂL, Gomes CFS, dos Santos M, do Carmo Silva M, Araujo JVGA. PROMETHEE-  
SAPEVO-M1 a Hybrid Modeling Proposal: Multicriteria Evaluation of Drones for Use in Naval  
Warfare. *Springer Proceedings in Mathematics & Statistics*. 1st ed., Cham: Springer; 2020, p. 381–  
93. [https://doi.org/10.1007/978-3-030-56920-4\\_31](https://doi.org/10.1007/978-3-030-56920-4_31).
- [7] Song Z, Yan T, Ge Y. Spatial equilibrium allocation of urban large public general hospitals based on the welfare maximization principle: A case study of Nanjing, China. *Sustainability* (Switzerland) 2018;10. <https://doi.org/10.3390/su10093024>.
- [8] R Core Team. R: A Language and Environment for Statistical Computing 2021.
- [9] Moreira MÂ, de Araújo Costa IP, Pereira MT, dos Santos M, Gomes CF, Muradas FM. PROMETHEE-SAPEVO-M1 a Hybrid Approach Based on Ordinal and Cardinal Inputs: Multi-Criteria Evaluation of Helicopters to Support Brazilian Navy Operations. *Algorithms* 2021;14. <https://doi.org/10.3390/a14050140>.
- [10] Ramachandran KM, Tsokos CP. Mathematical statistics with applications in R. Academic Press; 2020.
- [11] Zanjirani Farahani R, Hekmatfar M, editors. Facility Location. Heidelberg: Physica-Verlag HD; 2009. <https://doi.org/10.1007/978-3-7908-2151-2>.
- [12] Arrumdanay AC, Sari PP, Rahmadani P, Lubis AI. Web-Based Geographic Information System (GIS) in Determining Shortest Path of MSME Medan City Using Bellman-Ford Algorithm. *Journal of Physics: Conference Series*, vol. 1255, IOP Publishing; 2019, p. 12075.
- [13] Karive O, Hakimi S. The p-median problems. In: An algorithmic approach to network location problems. *SIAM Journal on Applied Mathematics* 1979;37:539–60.
- [14] Li Q, Zeng B, Savachkin A. Reliable facility location design under disruptions. *Computers and Operations Research* 2013;40:901–9. <https://doi.org/10.1016/j.cor.2012.11.012>.
- [15] Frederick Hillier, Lieberman G. Introduction to Operations Research. 11th ed. New York: McGraw-Hill Education; 2020.
- [16] Berkelaar M. Lpsolve 2020.
- [17] Cheng J, Karambelkar B, Xie Y, Wickham H. Leaflet 2021.

# Bibliometric Studies on Multi-Criteria Decision Analysis (MCDA) Applied in Personnel Selection

Igor Pinheiro de Araújo COSTA<sup>a,d,1</sup>, Marcio Pereira BASÍLIO<sup>b</sup>, Sérgio Mitihiro do Nascimento MAÊDA<sup>a,d</sup>, Marcus Vinícius Gonçalves RODRIGUES<sup>a</sup>, Miguel Ângelo Lellis MOREIRA<sup>a,d</sup>, Carlos Francisco Simões GOMES<sup>a</sup> and Marcos dos SANTOS<sup>c,d</sup>

<sup>a</sup>*Federal Fluminense University, Brazil*

<sup>b</sup>*Military Police of the State of Rio de Janeiro, Brazil*

<sup>c</sup>*Military Institute of Engineering, Brazil*

<sup>d</sup>*Naval Systems Analysis Center, Brazil*

**Abstract.** Personnel selection is increasingly proving to be an essential factor for the success of organizations. These issues almost universally involve multiple conflicting objectives, uncertainties, costs and benefits in the decision-making process. In this context, due to the characteristics of human resources problems, composed of several subjective evaluation criteria, such problems can be solved by applying Multiple Criteria Decision Aid (MCDA) methods. This paper aims to present a literature review on the main applications of MCDA in the personnel selection area, considering the tactical, operational and strategic spheres. The methodology includes a bibliometric study and a literature review of documents from the Scopus database. We identified the document type, year of publication, affiliation, authors, author's H-index, the field of knowledge, country and keyword clusters. The literature review allowed us to verify that the Fuzzy Logic is the most applied methodology in personnel selection problems with MCDA, due to its capacity to handle vague, imprecise and subjective data.

**Keywords.** Multi-criteria Decision Analysis (MCDA), Bibliometric Studies, Personnel Selection, Fuzzy Logic.

## 1. Introduction

Personnel selection is one of the problems that organizations are increasingly facing, as people represent one of the most important resources in companies [1]. According to [2], this selection is a significant task, as it aims to select the best candidates to perform the business functions.

In this sense, Multi-criteria Decision Aid (MCDA) describes methods that address multiple criteria to help individuals and groups make important decisions. [3]. MCDA is composed of a set of alternatives, at least two criteria and at least one Decision Maker (DM) [4].

---

<sup>1</sup> Corresponding Author, Igor Pinheiro de Araújo COSTA, Department of Production Engineering, Federal Fluminense University, Niterói, Brazil; E-mail: [costa\\_igor@id.uff.br](mailto:costa_igor@id.uff.br).

This research was financially supported by Naval Systems Analysis Center (CASNAV), Brazil.

The academic literature presents several applications of MCDA methods in personnel selection, in the tactical, operational and strategic spheres. However, there is a lack of studies on publishing trends in this important area for business and human resources management. Therefore, this research aims to fill this gap, presenting a literature review on the main applications of MCDA in the personnel selection problems.

Following the methodology presented by [5], we analyzed MCDA applications in personnel selection, to explore: the main authors, countries, journals and knowledge areas that publish papers on the subject; the number of works published; and the main keywords used in the papers.

This paper presents the results of bibliographic research on MCDA and personnel selection, providing a descriptive overview of the scientific production of both themes. A bibliometric study was performed on the Scopus database to achieve this goal.

## **2. Background**

In the personnel selection process, several attributes are usually evaluated such as analytical ability, innovation, leadership and personality. Therefore, Human Resources Management (HRM) plays a significant role in business administration [5].

For [5,6] personnel selection problems involve multiple candidates as alternatives, together with many complex criteria, which can be solved by MCDA methods. The multicriteria methods have been applied to support the decision-making process in recent complex problems, such as [7–13].

Nabeeh et al. [14] claim that the complexity of this type of problem is to establish the appropriate application to meet the company's requirements, which makes the process excessively subjective and complex. The environment of decision-making is surrounded by inconsistency and uncertainty [14].

For [13], without a proper and accurate method for HR problems, the performance of the project will be affected, and the MCDA is suggested to be a viable method for this kind of problem.

Among the main applications of MCDA in personnel selection, we highlight: a hybrid MCDA model for sniper selection [16]; competency frameworks for choosing the best information technology (IT) expert [6,16]; selection of a chief accounting officer, managing information assessed using both linguistic and numerical scales [17]; selection of best employees to a supply chain firm [18]; selection from security personnel [19]; selection of a candidate for the position of the sales manager [20].

The literature presents several applications combining MCDA methods to support the decision-making process in personnel selection problems, in most cases applying a method to obtain the weights of the criteria and another one to evaluate the alternatives, taking advantage of each method's characteristics.

Also, we observed that the fuzzy logic is routinely applied in personnel selection problems with MCDA. Given the great subjectivity and uncertainties related to personnel selection, it is justified by the fact that the fuzzy approaches handle vagueness and imprecision better [21]. These findings corroborate the impressions of several authors, such as [22] and [23], who applied fuzzy logic in conjunction with multicriteria methods to deal with data characterized by inconsistency and uncertainty in multicriteria problems.

### 3. Methodology

This study considered the Webibliomining model [24], chosen for integrating concepts from Bibliometrics, Webmetrics, Informetrics and Bibliomining. In this section, an adaptation of the model proposed by [24–26] was applied. The following strategy was tested in the Scopus database to find documents on applications of MCDA methods in personnel selection problems, linking both research themes:

- TITLE-ABS-KEY ((MCDA OR MCDM) AND (PERSONNEL SELECTION)).

The research was carried out in the Scopus database in May 2021 (Table 1). There were no limits of date, document type, or access type. In all, 63 studies were found with the themes analyzed.

**Table 1.** Search results in the SCOPUS database.

Document type	Number of articles	%
Journal article	52	82.5%
Conference paper	9	14.3%
Book Chapter	2	3.2%
Total	63	100.0%

A bibliometric study was developed to identify the year of publication, journals, clusters of keywords, authors (including the authors' H-index), affiliation, country/territory and fields of knowledge. VOSviewer software was used to analyze keyword clusters. It is a tool for creating maps, viewing and exploring [27]. According to [28], increasing publication rates and fragmented research streams make the use of bibliometric essential for scientific mapping.

### 4. Results and analysis

The first paper was published in 1994. From 1995 to 2003 and 2006 to 2009 there were no published articles. From 2010 to 2020, there was a significant increase in the number of articles published with fluctuations from 2 to 11 articles per year. In 2018, there was the highest number of publications (11 articles).

Table 2 shows the distribution of articles by journal, considering 2 or more works. The journals Expert Systems With Applications and Life Science presented the highest number of articles published in the area of personnel selection with MCDA applications (4 papers). There are also 9 journals with 2 articles each, and 26 journals with 1 published article, totaling 37. Therefore, it is noted that the articles are distributed by a wide variety of journals.

**Table 2.** Distribution of articles by journal.

Journal	Number of articles
Expert Systems With Applications	4
Life Science Journal	4
European Journal Of Operational Research	2
Applied Soft Computing Journal	2
Economic Computation And Economic Cybernetics Studies And Research	2
Human Factors And Ergonomics In Manufacturing	2
Informatica (Netherlands)	2
Journal Of Enterprise Information Management	2
Journal Of Multiple Valued Logic And Soft Computing	2
Scientific World Journal	2
Soft Computing	2

Analyzing the words can bring information and knowledge about a certain subject [29]. This research used VOSviewer software to analyze keywords, including author and index keywords. The keyword "personnel selection" has the highest number of occurrences, followed by "decision making". The terms most related to personnel problems are "risk assessment", "human resources" and "enterprise resource management". Among the MCDA methods, the Analytic Hierarchy Process (AHP), Technique for Order Preferences by Similarity To Ideal Solution (TOPSIS), Additive Ratio Assessment (ARAS) and *Elimination Et Choix Traduisant la Réalité* (ELECTRE) stand out in the analysis of keywords.

The first applications dealt with HR management, while the most recent ones seem to indicate a trend of research more focused on risk assessment, with qualitative approaches involving linguistic variables and fuzzy logic. Table 3 shows the distribution of articles per author, in descending order according to the number of published papers and H-Index, considering 3 or more published articles. The H-index is defined as the number of articles with citation numbers greater than or equal to H, as a useful index to characterize the scientific production of a researcher [30]. There are 2 authors with 6 articles: Stanujkic, D. (H = 19) and Karabasevic, D. (H = 13). 2 authors published 5 articles each, with emphasis on the author Zavadskas, E.K., who has an H-index = 77. Besides, there is an author with 4 articles, 3 authors with 3 papers, 14 researchers with 2 articles and 116 authors with one article each, totaling 138 authors.

**Table 3.** Distribution of articles by author.

Author	Articles	H-Index
Stanujkic, D.	6	19
Karabasevic, D.	6	13
Zavadskas, E.K.	5	77
Turskis, Z.	5	54
El-Santawy, M.F.	4	4
Aliguliyev, R.	3	20
Ahmed, A.N.	3	4
Yusifov, F.	3	2

Table 4 shows the institutions that have published 2 or more papers. The institution *Vilniaus Gedimino Technikos Universitetas* stands out with 11 documents, followed by the University of Belgrade, with 5 articles. Also, there are 2 institutions with 3 and 4 articles each, 5 universities with 2 papers and 72 institutions with 1 article each, totaling 83.

**Table 4.** Distribution of articles by institution.

Institution	Articles
<i>Vilniaus Gedimino Technikos Universitetas</i>	11
University of Belgrade	5
Cairo University	4
Azerbaijan National Academy of Sciences	4
University Business Academy in Novi Sad	3
<i>Galatasaray Üniversitesi</i>	3
Zagazig University	2
University of Tehran	2
The University of New Mexico – Gallup	2
Ming Chuan University	2
Mykolas Romeris University	2

Table 5 shows the distribution of articles by country or territory. The search found 21 countries. Lithuania ranks first with 12 articles, followed by Turkey with 11

documents. Asia, with 9 countries, has 39 articles (62% of the total). Europe is represented by 8 countries with 28 articles published (44% of the total). Africa, represented by 2 countries, has 7 published articles (11%). The American continent, with 1 country, has 3 articles (5% of the total). Oceania, through New Zealand, has 1 article (2% of the total).

**Table 5.** Distribution of articles by country or territory.

Country/territory	Articles
Lithuania	12
Turkey	11
Taiwan	9
Egypt	6
India	6
Serbia	6
Azerbaijan	5
Iran	4
China	3
United States	3
Vietnam	3

Table 6 shows the distribution of articles by field of knowledge: Computer science (28.7%), Engineering (14.7%), Mathematics (12.4%), Business Management (9.3%) and Decision Sciences (8.5%) stand out compared to other areas, with approximately 74% of the total articles.

**Table 6.** Distribution of articles by area of knowledge.

Subject area	Articles (%)
Computer Science	37 (28.7%)
Engineering	19 (14.7%)
Mathematics	16 (12.4%)
Business, Management and Accounting	12 (9.3%)
Decision Sciences	11 (8.5%)
Economics, Econometrics and Finance	7 (5.4%)
Biochemistry, Genetics and Molecular Biology	6 (4.7%)
Social Sciences	6 (4.7%)

## 5. Conclusions

The bibliometric study provided a descriptive overview of scientific production on applications of MCDA methods in personnel selection problems. The literature review, although not exhaustive, showed several concepts, paradigms, steps and applications in different personnel problems. The papers are divided by several journals, and there is not one that can be pointed out as the greatest reference in personnel selection with MCDA.

Regarding the authors, no one can be considered as the largest reference in the area. The distribution by field of knowledge showed that the articles are spread over several areas, illustrating the multidisciplinarity of the themes under study.

The analysis of the keywords allowed us to verify that, among the MCDA methods, four of them stand out: AHP, TOPSIS, ARAS and ELECTRE. Also, through the temporal analysis of keywords, we verified that the most recent applications seem to indicate a trend of research more focused on risk assessment, with qualitative approaches involving linguistic variables and fuzzy logic. This trend is probably

justified by the capacity to deal with imprecise, vague, and subjective data that Fuzzy Logic offers the DM.

Because of the above, this research fulfilled the gap in the literature, presenting the state of the art of applications in MCDA methods in a field that is so important for the success of companies - personnel selection. As future work, we suggest more detailed studies on the competencies considered most important in the process of choosing employees in the most diverse fields of human activity.

## References

- [1]. Ulutaş A, Popovic G, Stanujkic D, Karabasevic D, Zavadskas EK, Turskis Z. A new hybrid mcdm model for personnel selection based on a novel grey piprecia and grey OCRA methods. Mathematics [Internet]. 2020;8(10):1–14. Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85092918383&doi=10.3390%2Fmath8101698&partnerID=40&md5=fd6228a310ade407ef6f60dff9efc0c>
- [2]. Abdel-Basset M, Gamal A, Son LH, Smarandache F. A bipolar neutrosophic multi criteria decision making framework for professional selection. Appl Sci [Internet]. 2020;10(4). Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85081200450&doi=10.3390%2Fapp10041202&partnerID=40&md5=18047ba29b237adf782c23417b73db70>
- [3]. Belton V, Stewart T. Multiple criteria decision analysis: an integrated approach. Springer Science & Business Media; 2002.
- [4]. Greco S, Figueira J, Ehrgott M. Multiple Criteria Decision Analysis: State of art surveys. Vol. 37. Springer; 2016.
- [5]. Ijadi Maghsoodi A, Riahi D, Herrera-Viedma E, Zavadskas EK. An integrated parallel big data decision support tool using the W-CLUS-MCDA: A multi-scenario personnel assessment. Knowledge-Based Syst [Internet]. 2020;195. Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85082191498&doi=10.1016%2Fj.knosys.2020.105749&partnerID=40&md5=0080ce509da1f9b4e6adb528954fad7a>
- [6]. Raj Mishra A, Sisodia G, Raj Pardasani K, Sharma K. Multi-criteria it personnel selection on intuitionistic fuzzy information measures and aras methodology. Iran J Fuzzy Syst [Internet]. 2020;17(4):55–68. Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85085562628&doi=10.22111%2Fijfs.2020.5406&partnerID=40&md5=b85b97852e9c2fc10c8d83fe59298565>
- [7]. Costa IP de A, Sanseverino AM, Barcelos MR dos S, Belderrain MCN, Gomes CFS, Santos M dos. Choosing flying hospitals in the fight against the COVID-19 pandemic: structuring and modeling a complex problem using the VFT and ELECTRE-MOr methods. IEEE Lat Am Trans. 2021;19(6):1099–106.
- [8]. Moreira MÂL, Costa IP de A, Pereira MT, dos Santos M, Gomes CFS, Muradas FM. PROMETHEE- SAPEVO-M1 a Hybrid Approach Based on Ordinal and Cardinal Inputs: Multi-Criteria Evaluation of Helicopters to Support Brazilian Navy Operations. Algorithms [Internet]. 2021 Apr 27;14(5):140. Available from: <https://www.mdpi.com/1999-4893/14/5/140>
- [9]. Oliveira AS, Gomes CFS, Clarkson CT, Sanseverino AM, Barcelos MRS, Costa IPA, et al. Multiple Criteria Decision Making and Prospective Scenarios Model for Selection of Companies to Be Incubated. Algorithms [Internet]. 2021 Mar 30;14(111). Available from: <https://www.mdpi.com/1999-4893/14/4/111>
- [10]. Maêda; SM do N, Costa IP de A, Castro Junior MAP, Fávero LP, Costa AP de A, Corriça JV de P, et al. Multi-criteria analysis applied to aircraft selection by Brazilian Navy. Production. 2021;31:1–13.
- [11]. Santos M dos; Costa IP de A, Gomes CFS. Multicriteria decision-making in the selection of warships: a new approach to the AHP method. Int J Anal Hierarchy Process [Internet]. 2021 May 19;13(1). Available from: <https://ijahp.org/index.php/IJAHP/article/view/833>
- [12]. de Almeida IDP, Corriça JV de P, Costa AP de A, Costa IP de A, Maêda SM do N, Gomes CFS, et al. Study of the Location of a Second Fleet for the Brazilian Navy: Structuring and Mathematical Modeling Using SAPEVO-M and VIKOR Methods. ICPR-Americas 2020 Commun Comput Inf Sci [Internet]. 2021;1408:113–24. Available from: [https://link.springer.com/10.1007/978-3-030-76310-7\\_9](https://link.springer.com/10.1007/978-3-030-76310-7_9)

- [13]. Costa IP de A, Maêda SM do N, Teixeira LFH de S de B, Gomes CFS, Santos M dos. Choosing a hospital assistance ship to fight the Covid-19 pandemic. *Rev Saude Publica* [Internet]. 2020;54. Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85090141296&doi=10.11606%2FS1518-8787.2020054002792&partnerID=40&md5=90355a4a86a1b09b1add8956ace15019>
- [14]. Nabeeh NA, Smarandache F, Abdel-Basset M, El-Ghareeb HA, Aboelfetouh A. An Integrated Neutrosophic-TOPSIS Approach and Its Application to Personnel Selection: A New Trend in Brain Processing and Analysis. *IEEE Access* [Internet]. 2019;7:29734–44. Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85064717679&doi=10.1109%2FACCESS.2019.2899841&partnerID=40&md5=c8d63c0f9059bb4f005c3fcfe5841ef9>
- [15]. Cheng EWL, Li H. Contractor selection using the analytic network process. *Constr Manag Econ* [Internet]. 2004;22(10):1021–32. Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-11944249373&doi=10.1080%2F0144619042000202852&partnerID=40&md5=fac75ca73bea9d7150126682cf3eac5d>
- [16]. Kabak M, Burmaoğlu S, Kazançoğlu Y. A fuzzy hybrid MCDM approach for professional selection. *Expert Syst Appl* [Internet]. 2012;39(3):3516–25. Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-80255127315&doi=10.1016%2Feswa.2011.09.042&partnerID=40&md5=19e4bb9448212f89daf636c6cf2602d>
- [17]. Keršulienė V, Turskis Z. A hybrid linguistic fuzzy multiple criteria group selection of a chief accounting officer. *J Bus Econ Manag* [Internet]. 2014;15(2):232–52. Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84899869374&doi=10.3846%2F16111699.2014.903201&partnerID=40&md5=fc7dd4831783fee4077ea06db0fc7602>
- [18]. Dwivedi P, Chaturvedi V, Vashist JK. Efficient team formation from pool of talent: comparing AHP-LP and TOPSIS-LP approach. *J Enterp Inf Manag* [Internet]. 2020;33(5):1293–318. Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85084979643&doi=10.1108%2FJEIM-09-2019-0283&partnerID=40&md5=c9ca04552dd3888fc77bb4e80b6c182>
- [19]. Krylovas A, Zavadskas EK, Kosareva N, Dadelo S. New KEMIRA method for determining criteria priority and weights in solving MCDM problem. *Int J Inf Technol Decis Mak* [Internet]. 2014;13(6):1119–33. Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84929300521&doi=10.1142%2F80219622014500825&partnerID=40&md5=61a8e591d6f9b1eb58067da73b64a5a9>
- [20]. Karabasevic D, Zavadskas EK, Turskis Z, Stanujkic D. The Framework for the Selection of Personnel Based on the SWARA and ARAS Methods Under Uncertainties. *Inform* [Internet]. 2016;27(1):49–65.
- [21]. Krishankumar R, Premaladha J, Ravichandran KS, Sekar KR, Manikandan R, Gao XZ. A novel extension to VIKOR method under intuitionistic fuzzy context for solving personnel selection problem. *Soft Comput* [Internet]. 2020;24(2):1063–81. Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85064160864&doi=10.1007%2Fs00500-019-03943-2&partnerID=40&md5=1406f69b9844add7b5578e50d4382267>
- [22]. Baležentis T, Zeng S. Group multi-criteria decision making based upon interval-valued fuzzy numbers: An extension of the MULTIMOORA method. *Expert Syst Appl* [Internet]. 2013;40(2):543–50.
- [23]. Kilic HS, Demirci AE, Delen D. An integrated decision analysis methodology based on IF-DEMATEL and IF-ELECTRE for personnel selection. *Decis Support Syst* [Internet]. 2020;137.
- [24]. Costa HG. Model for webbibliomining: proposal and application. *Rev FAE*. 2010;115–26.
- [25]. De Barros MD, Salles CAL, Gomes CFS, Da Silva RA, Costa HG. Mapping of the scientific production on the ITIL application published in the national and international literature. In: R. C, Y. S, P. W, E. H-V, L.F.A.M. G, Y. S, editors. Universidade Federal Fluminense, 156 - Sao Domingos Rua Passos da Patria, Niteroi - RJ, 24210-240, Brazil: Elsevier B.V.; 2015. p. 102–11.
- [26]. Da Silva GB, Costa HG, De Barros MD. Entrepreneurship in engineering education: A literature review. *Int J Eng Educ*. 2015;31(6):1701–10.
- [27]. Van Eck NJ, Waltman L. Manual for VOSviewer version 1.6. 8. CWTS Meaningful Metrics Univ Leiden. 2018;
- [28]. Aria M, Cuccurullo C. bibliometrix: An R-tool for comprehensive science mapping analysis. *J Informetr*. 2017;11(4):959–75.
- [29]. Ishikiriyama CS, Miro D, Gomes CFS. Text Mining Business Intelligence: a small sample of what words can say. *Procedia Comput Sci*. 2015;55:261–7.
- [30]. Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci*. 2005;102(46):16569–72.

# Multicriteria Analysis in Additive Manufacturing: An ELECTRE-MOr Based Approach

Paula DRUMOND<sup>a</sup>, Marcio Pereira BASÍLIO<sup>b</sup>, Igor Pinheiro de Araújo COSTA<sup>c,1</sup>, Daniel Augusto de Moura PEREIRA<sup>d</sup>, Carlos Francisco Simões GOMES<sup>c</sup> and Marcos dos SANTOS<sup>a</sup>

<sup>a</sup>*Military Institute of Engineering, Brazil*

<sup>b</sup>*Military Police of the State of Rio de Janeiro, Brazil*

<sup>c</sup>*Federal Fluminense University, Brazil*

<sup>d</sup>*Federal University of Campina Grande, Brazil*

**Abstract.** 3D printing technologies define the essence of Additive Manufacturing and make possible the agile production of customized parts from different materials, with lower unit cost and waste generation. Currently, one of the most widespread 3D printer technologies is the Fused Deposition Modeling (FDM) type, which is the object of this paper. The choice of 3D printing equipment depends on the alignment of the purpose of use and technical knowledge to consider certain requirements. Therefore, this choice can be time-consuming and/or imprecise. In this sense, this work aimed to classify FDM-type 3D printer models by applying the ELECTRE-MOr method, a Multi-criteria Decision Aiding (MCDA) method. As a result, based on a categorization between classes, the FABER 10 alternative was the only one that presented class A performance in all evaluated scenarios, based on criteria defined by the experts consulted in this study.

**Keywords.** Multi-criteria Decision Aiding (MCDA), ELECTRE-MOr, Additive Manufacturing, Decision Making.

## 1. Introduction

Additive Manufacturing (AM) is the technology capable of transforming a complex geometry model, elaborated through a computational system, into a product without the need for a long production planning process. The manufacture of parts happens with the overlap of material layers, added one by one in a three-dimensional printing area, after software configurations [1]. According to [2], AM can be considered as the manufacturing paradigm of the future due to its ability to meet changes in capacity and functionality quickly and efficiently.

AM has gained wide diffusion in a number of applications based on the layer manufacturing technique in recent years [3]. However, it is a challenge to select an

---

<sup>1</sup> Corresponding Author, Igor Pinheiro de Araújo COSTA, Department of Production Engineering, Federal Fluminense University, Niterói, Brazil; E-mail: costa\_igor@id.uff.br.

This research was financially supported by Military Police of the State of Rio de Janeiro.

Appropriate AM process for a particular customer or company. This is because the result may vary due not only to different materials and printers, but also to different parameters and post-processes [4].

According to [5], selecting an appropriate process or machine to manufacture an end-use product is an important issue in AM design. One of the many types of approaches to process selection is based on Multi-criteria Decision Aiding (MCDA). In this context, MCDA consists of a set of formal approaches that explicitly seek to take into account multiple criteria to help stakeholders and groups explore important decisions [6]. These decisions generally involve various conflicting objectives, cloudy types of non-repeatable uncertainties, accumulated costs and benefits for various individuals, companies, groups and other organizations [7].

Despite the diversity of multi-criteria approaches, methods and techniques, the basic components of MCDA are a finite or infinite set of actions (alternatives, solutions, courses of action, etc.), at least two criteria and at least one Decision Maker (DM). Given these basic elements, MCDA is an activity that assists in decision making, especially in terms of choice, classification or ordering of actions [8].

This paper aims proposes, through the application of the ELECTRE-MOr MCDA method, an algorithm capable of supporting the strategic process to classify 3D printing equipment for AM. The axiomatic model of the ELECTRE-MOr method allows the evaluation and obtaining of the weights of the criteria, through qualitative pairwise analysis by specialists in AM, and the distribution of alternatives into predefined classes. The method was chosen because it distributes alternatives in classes, which allows the DM to select one or more options that are allocated to the highest classes.

This research is divided into six sections. After this introduction, section 2 presents the literature review; section 3 structures the problem and the methodology, while section 4 deals with the main characteristics of the ELECTRE-MOr method; section 5 demonstrates its application and results; finally, section 6 concludes this research.

## 2. Literature Review

With Industry 4.0 trends covering physical and virtual media, AM represents a way for the realization of complex geometries designed on computers [9]. Widely known as 3D Printing, AM has been spreading around the world due to the possibility of developing customized products quickly and with minimal material waste, which positively impacts the production chain [1].

In this context, the Fused Deposition Modeling (FDM) printer uses heat to melt a thermoplastic filament and then generates 3D item from digital designs [10], presenting an economic process, mainly in the construction of items with complex geometry or almost any shape [11].

The academic literature presents some MCDA methods' applications in problems related to AM, such as: evaluation of the MA machines capacity, based on the ability to produce a standard component [12]; a proposed hybrid modeling to help users in the AM options selection process [4].

Khamhong et al. [13] used the Fuzzy Analytic Hierarchy Process (Fuzzy AHP) methodology to analyze criteria weights for factors related to the selection of 3D printers. Raigar et al. [14] proposed a hybrid multicriteria methodology to select an appropriate AM process from available processes. The Best Worst Method (BWM) was applied to

determine the optimal weights of the criteria, and the Proximity Indexed Value (PIV) method was used to classify the available AM processes.

According to [15], the costs of the machine and the material are the significant parameters, which play an important role in estimating the cost of the AM. The authors applied the BWM method to select the appropriate material for an AM project, aiming to reduce the costs inherent to the project.

According to [16], although the options are plentiful, selecting an appropriate choice of AM is not a trivial task. The authors applied the Combinative Distance-Based Assessment (CODAS) method under the Pythagorean Fuzzy (PF) to assist in the selection of the 3D printing technology best suited to the needs of the supply chain in a given company.

The literature review, although not exhaustive, revealed several MCDA applications to support the decision-making process in AM-related problems. However, it was verified the lack of models for sorting alternatives in classes, as is the proposal of this paper.

The modeling presented in this research distributes the alternatives in predefined categories, presenting ordinal entry of weights, multiple DMs and evaluation of quantitative and qualitative criteria. Furthermore, the ELECTRE-MOR method presents two ways to obtain the lower bounds of predefined classes, resulting in two alternative distributions. This feature provides greater transparency and robustness to the decision-making process.

### **3. Methodology**

Considering a real decision-making problem, uncertainty is intrinsic. As suggested by [17], the methods used as decision-aid should enable an integrated algorithm, enabling the evaluation of qualitative and quantitative data. That is, tools should be able to structure and analyze variables in situations where it is not possible to define a precise numeric entry. In the decision analysis, the presence of uncertainty regarding the evaluation of a problem is recurrent, considering that the information obtained may be lacking in complete or certainty data [18].

During the analysis of the data referring to the case study of this paper, we found that, among the diverse available MCDA tools, the ELECTRE-MOr method has good adherence to the problem, because it distributes the alternatives in predefined classes and allows obtaining the criteria weights by ordinal entries, considering the opinion of multiple DMs.

Therefore, this work aims to propose an action plan for classification of FDM 3D printers, based on the application of the ELECTRE-MOr method, which facilitates the expression of preference relationships for DMs, contributing to transparent and reliable decision making.

According to the classification proposed by [19], this research can be characterized as mixed qualitative-quantitative, combining case study and mathematical modeling [20]. AM is the object of research, as previously presented in Sections 1 and 2. The background and details about the case study are presented in Sections 4 and 5, respectively.

#### 4. The ELECTRE-MOr method

The MCDA methods are very useful to support the decision-making process in several situations, because they consider value judgments and not only technical issues, to evaluate alternatives to solve real problems, presenting a highly multidisciplinarity [21]. These methods have been employed to support the decision-making process in several recent complex problems, as presented in [22–26].

The ELECTRE MOr, proposed by [27], is a multicriteria sorting method with ordinal weight input that includes multiple DMs and distributes the alternatives into pre-defined categories. The ELECTRE MOr procedures are developed in two stages. To obtain the weights and evaluate qualitative criteria, the ELECTRE-MOr method uses an adaptation of the SAPEVO method [28,29], transforming ordinal preferences of criteria into a vector of criteria weights and integrating the vector criteria of different DMs.

According to [27], ELECTRE-MOr has the following features:

- The elicitation of weights of the criteria by an ordinal form, since this is not an easy task for a DM, because it requires establishing a precise numerical value for such parameters as the importance coefficients of criteria;
- The distribution in classes is carried out through two procedures:
  1. Optimistic: it consists of comparing the alternative successively to alternative b, from the last profile (category, class);
  2. Pessimistic: it consists of comparing the alternative a successively to the alternative b, starting from the first profile (category, class), which is the most demanding classification.
- Two ways of obtaining the lower limits of the classes (bh and bn), which provides 4 different sorts (2 optimistic and 2 pessimistic), allowing a more robust and reliable sensitivity analysis of the results.

#### 5. Case Study

Given the numerous options of FDM 3D printers found in the market, we consulted three specialists (DMs) in the AM field, with extensive experience and knowledge in 3D printers, to designate the alternative components of this research. For possible alternatives, printers that have similarities between construction pattern (equipment and finished product) and price were considered, so that there was no inequality between the selected alternatives. Thus, it was concluded that eight printers would be evaluated: ENDER 3, ENDER 3 PRO, DAVINCI PRO 1.0, GI3, S2, PRIME ONE V3, GRABER I3 and FABER 10.

Also, the DMs defined six relevant criteria that relate to the choice of equipment and that can influence the quality of the 3D printed parts (Table 1):

**Table 1.** Establishment of outranking relationships.

Acronym	Criteria	Monotonicity
C <sub>1</sub>	Print speed (mm/s)	Profit
C <sub>2</sub>	Printing area (mm <sup>2</sup> )	Profit
C <sub>3</sub>	Price (R\$)	Cost
C <sub>4</sub>	Layer resolution (mm)	Cost
C <sub>5</sub>	Warranty (Months)	Profit
C <sub>6</sub>	Energy consumption (w)	Cost

The monotonicity of the criteria was determined, that is, whether they are cost or benefit criteria, in order to minimize or maximize them, respectively, when incorporated into the ELECTRE-MOr method. In order to obtain the weights of the criteria (through pairwise comparisons between them), interviews were conducted through video conference with the three specialists together to establish the degrees by consensus.

### 5.1. Evaluation of alternatives

Table 2 shows the performance matrix, with the alternatives being evaluated in the light of each criterion. To obtain data referring to criteria  $C_1$ ,  $C_2$ ,  $C_4$ ,  $C_5$  and  $C_6$ , we considered parameters contained in the equipment's technical manuals. To obtain the price ( $C_3$ ), a market research was carried out.

**Table 2.** Performance matrix.

Alternatives / Criteria	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$
ENDER 3	100	48.400	-2.199	-0.050	6	-340
ENDER 3 PRO	100	48.40	-2.599	-0.050	6	-340
DAVINCI PRO 1.0	120	40.000	-10.000	-0.075	12	-200
GI3	200	40.000	-3.100	-0.050	12	-360
S2	150	40.000	-5.550	-0.050	12	-350
PRIME ONE V3	75	40.000	-2.350	-0.080	3	-360
GRABER I3	80	40.000	-1.950	-0.050	12	-360
FABER 10	200	90.000	-3.685	-0.070	12	-270
$bh_3$	169	77.500	-3.963	-0.058	10	-240
$bh_2$	138	65.000	-5.975	-0.065	8	-280
$bh_1$	106	52.500	-7.988	-0.073	5	-320
$bn_3$	150	48.400	-2.199	-0.070	12	-270
$bn_2$	100	40.000	-2.599	-0.075	6	-350
$bn_1$	80	40.000	-3.685	-0.080	3	-360
q	15	2.000	100	0.025	6	15
p	20	4.000	500	0.050	12	50
v	25	10.000	10.000	0.075	10	200
Weights of criteria	0.92	1.21	2.00	1.29	0.72	0.45

It is observed that the criterion with the highest weight after analysis of the three specialists was the price, followed by the resolution of the layer and printing area. The energy consumption criterion was considered as the least important by analysts.

The monotonic cost criteria were represented with negative values in the performance matrix. The Weak Preference (q), Strict Preference (p) and Veto (v) thresholds were defined by consensus among the experts. In addition, it was established that the eight alternatives would be distributed into four classes.

### 5.2. Results obtained

After applying all the steps of ELECTRE-MOr, optimistic and pessimistic classifications were obtained for the two forms of distribution ( $bh$  and  $bn$ ). In consensus among the authors and specialists, a cutting level  $\lambda$  of 0.6 was established, because it presented a good discrimination in the classification of alternatives. Thus, the distribution of alternatives in the four predefined classes is obtained (Table 3).

**Table 3.** Results obtained after application of the ELECTRE-MOr method.

$\lambda = 0,6$ Alternatives	bh		bn	
	Pessimist	Optimistic	Pessimist	Optimistic
ENDER 2	C	C	B	B
ENDER 3 PRO	C	C	B	B

DAVINCI PRO 1.0	D	D	B	B
GI3	D	A	A	A
S2	D	C	A	A
PRIME ONE V3	D	D	C	C
GRABER I3	D	D	B	B
FABER 10	A	A	A	A

In view of the classifications obtained through the method, it was observed that the FABER 10 alternative was unique with the classification corresponding to Class A in both scenarios (pessimistic and optimistic) within the two distributions (bh and bn). Therefore, this is the most suitable 3D printer to be purchased in the face of the conditions addressed.

In the analysis of bh, it is noted that only FABER 10 achieved the best classification in both scenarios. Analyzing the bn procedure, it is perceived that the alternatives GI3, S2 also reached the highest class, but GI3 obtained a bad result in the pessimistic scenario of bh and S2 had poor performance in both bh scenarios.

Among the reasons that led FABER 10 to be the best alternative, it is possible to observe the good overall performance in the criteria, highlighting the large printing area (third criterion of higher weight) of 90,000 mm<sup>2</sup> and the printing speed of 200 mm/s, which justifies the good performance of the alternative in the proposed analysis.

## 6. Conclusions

This article aimed to support the classification process of 3D printers of the FDM type, in order to evaluate the alternatives selected in the light of established criteria to reach a result based on a multicriteria model. The application of the ELECTRE-MOr method considered the evaluation of three different specialists, based on the criteria in question and the weights derived from their evaluations, having interfered on the final result.

We highlight that the method proved to be efficient for the proposed analysis, allowing the entry of ordinal weights, considering the opinion of multiple DMs, both in obtaining the weights, as well as in elicitation of the preference and veto thresholds. Besides, the presentation of two pessimistic and optimistic ordinations allowed us to verify the behavior of alternatives in various scenarios, with the possibility of choosing several parameters of analysis, either by considering the highest number of occurrences of classifications, as was done in this paper, or by considering only one of the two forms of distribution. As a suggestion for future work, it is proposed that other relevant and/or alternative criteria be included, modifying the cutting level, and that 3D printers of other types be analyzed.

## References

- [1]. GIBSON I., ROSEN DW., STUCKER B. Additive Manufacturing Technologies: 3D Printing, rapid prototyping and direct digital manufacturing. New York: Springer; 2015.
- [2]. Sonar HC, Khanzode V V, Akarte MM. Ranking of Additive Manufacturing Implementation Factors using Analytic Hierarchy Process (AHP). J Inst Eng Ser C [Internet]. 2021;102(2):421–6.
- [3]. Ransikarbum K, Khamhong P. Integrated Fuzzy Analytic Hierarchy Process and Technique for Order of Preference by Similarity to Ideal Solution for Additive Manufacturing Printer Selection. J Mater Eng Perform [Internet]. 2021.
- [4]. Wang Y, Zhong RY, Xu X. A decision support system for additive manufacturing process selection using a hybrid multiple criteria decision-making method. Rapid Prototyp J [Internet]. 2018;24(9):1544–53.
- [5]. Qin Y, Qi Q, Scott PJ, Jiang X. An additive manufacturing process selection approach based on fuzzy

- Archimedean weighted power Bonferroni aggregation operators. *Robot Comput Integr Manuf.* 2020;64.
- [6]. Belton V, Stewart T. Multiple criteria decision analysis: an integrated approach. Springer Science & Business Media; 2002.
- [7]. Keeney RL, Raiffa H, Meyer RF. Decisions with multiple objectives: preferences and value trade-offs. Cambridge university press; 1993.
- [8]. Greco S, Figueira J, Ehrgott M. Multiple Criteria Decision Analysis: State of art surveys. Vol. 37. Springer; 2016.
- [9]. Oh Y, Witherell P, Lu Y, Srock T. Nesting and scheduling problems for additive manufacturing: A taxonomy and review. *Addit Manuf* [Internet]. 2020;36.
- [10].Joo Y, Shin I, Ham G, Abuzar SM, Hyun S-M, Hwang S-J. The advent of a novel manufacturing technology in pharmaceuticals: Superiority of fused deposition modeling 3D printer. *J Pharm Investig.* 2020;50(2):131–45.
- [11].Tan DK, Maniruzzaman M, Nokhodchi A. Advanced pharmaceutical applications of hot-melt extrusion coupled with fused deposition modelling (FDM) 3D printing for personalised drug delivery. *Pharmaceutics.* 2018;10(4):203.
- [12].Roberson DA, Espalin D, Wicker RB. 3D printer selection: A decision-making evaluation and ranking model. *Virtual Phys Prototyp* [Internet]. 2013;8(3):201–12.
- [13].Khamhong P, Yingwiwatanapong C, Ransikarbum K. Fuzzy Analytic Hierarchy Process (AHP)-based Criteria Analysis for 3D Printer Selection in Additive Manufacturing. In: RI2C 2019 - 2019 Research, Invention, and Innovation Congress [Internet]. 2019.
- [14].Raigar J, Sharma VS, Srivastava S, Chand R, Singh J. A decision support system for the selection of an additive manufacturing process using a new hybrid MCDM technique. *Sadhana - Acad Proc Eng Sci* [Internet]. 2020;45(1).
- [15].Palanisamy M, Pugalendhi A, Ranganathan R. Selection of suitable additive manufacturing machine and materials through best-worst method (BWM). *Int J Adv Manuf Technol.* 2020;107(5–6):2345–62.
- [16].Büyüközkan G, Göçer F. Assessment of additive manufacturing technology by pythagorean fuzzy CODAS [Internet]. Vol. 1029, Advances in Intelligent Systems and Computing. 2020. p. 959–68.
- [17].Malloy TF, Zaunbrecher VM, Batteate CM, Blake A, Carroll WF, Corbett CJ, et al. Advancing alternative analysis: Integration of decision science. *Environ Health Perspect.* 2017;125(6):1–12.
- [18].Dong Y, Liu Y, Liang H, Chiclana F, Herrera-Viedma E. Strategic weight manipulation in multiple attribute decision making. *Omega (United Kingdom).* 2018;75:1339–51.
- [19].Creswell JW, Creswell JD. Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications; 2017.
- [20].Bertrand JWM, Fransoo JC. Operations management research methodologies using quantitative modeling. *Int J Oper Prod Manag.* 2002;22(2):241–64.
- [21].Santos M dos, Quintal RS, Paixão AC da, Gomes CFS. Simulation of Operation of an Integrated Information for Emergency Pre-Hospital Care in Rio de Janeiro Municipality. *Procedia Comput Sci* [Internet]. 2015;55:931–8.
- [22].Costa IP de A, Maëda SM do N, Teixeira LFH de S de B, Gomes CFS, Santos M dos. Choosing a hospital assistance ship to fight the Covid-19 pandemic. *Rev Saude Publica* [Internet]. 2020;54.
- [23].Tenório FM, dos Santos M, Gomes CFS, Araujo J de C. Navy Warship Selection and Multicriteria Analysis: The THOR Method Supporting Decision Making. In: International Joint conference on Industrial Engineering and Operations Management [Internet]. Springer; 2020. p. 27–39.
- [24].Moreira MÂL, Costa IP de A, Pereira MT, dos Santos M, Gomes CFS, Muradas FM. PROMETHEE- SAPEVO-M1 a Hybrid Approach Based on Ordinal and Cardinal Inputs: Multi-Criteria Evaluation of Helicopters to Support Brazilian Navy Operations. *Algorithms* [Internet]. 2021 Apr 27;14(5):140.
- [25].Oliveira AS, Gomes CFS, Clarkson CT, Sanseverino AM, Barcelos MRS, Costa IPA, et al. Multiple Criteria Decision Making and Prospective Scenarios Model for Selection of Companies to Be Incubated. *Algorithms* [Internet]. 2021 Mar 30;14(111).
- [26].Santos M dos, Costa IP de A, Gomes CFS. Multicriteria decision-making in the selection of warships: a new approach to the AHP method. *Int J Anal Hierarchy Process* [Internet]. 2021 May 19;13(1).
- [27].Costa IP de A, Sanseverino AM, Barcelos MR dos S, Belderrain MCN, Gomes CFS, Santos M dos. Choosing flying hospitals in the fight against the COVID-19 pandemic: structuring and modeling a complex problem using the VFT and ELECTRE-MOr methods. *IEEE Lat Am Trans* [Internet]. 2021 Jun;19(6):1099–106.
- [28].Gomes CFS, Santos M dos, Teixeira LFH de S de B, Sanseverino AM, Barcelos MR dos S. SAPEVO-M: a group multicriteria ordinal ranking method. *Pesqui Operacional* [Internet]. 2020;40.
- [29].Gomes L, Mury A-R, Gomes CFS. Multicriteria ranking with ordinal data. *Syst Anal.* 1997;27(2):139–46.

# Blockchain Digital Test Certificates for COVID-19

Ioannis Karamitsos<sup>a,1</sup> and Maria Papadaki<sup>b</sup>

<sup>a</sup>*Rochester Institute of Technology, Dubai, UAE*

<sup>b</sup>*British University in Dubai, Dubai, UAE*

**Abstract.** COVID-19 is a pandemic outbreak for each country worldwide. Each government needs to monitor every citizen and the COVID-19 test becomes an essential evidence for people who are travelling. This gives rise to the necessity of disruptive technologies such as Blockchain. In this paper, we provide an overview of the Hyperledger and Ethereum platforms and present how healthcare organizations can control and monitor digital health test certificates with citizens or other stakeholders. We also present a smart contract structure and implementation for COVID-19 test certificates in both blockchain platforms.

**Keywords.** Blockchain, COVID-19, Ethereum, Hyperledger, Health Certificate, Smart Contract

## 1. Introduction

Coronavirus Disease (COVID-19) [1] is an outbreak pandemic infectious disease which has appeared in 2019 from a Chinese city, named Wuhan. In few months, the COVID-19 moved from being an epidemic to a pandemic infectious disease. Infectious diseases are diseases that spread rapidly among large numbers of people within a short period of time, and they can be caused by several organisms such as viruses, bacteria or parasites. Naturally, those organisms live in human bodies harmlessly, but due to certain factors or drivers, they become harmful to the body. The transmission of infectious diseases can occur by humans, insects, or animals, and each infectious disease differs in terms of the causing factors, symptoms, the transmission approach and prevention control.

According to the World Health Organisation (WHO) [2], 14 million confirmed cases, including 607781 deaths globally, have been reported. Due to the high number of infected people, many countries locked down all the activities affecting negatively the economy and the organisations. The main purpose of the lockdown was to limit the mobility of the people in order to keep the dispersion of the disease low. On the other hand, governments are now phasing the next day for the economy and reshaping their countermeasures in our daily activities. One of the first releases was the mobility of the people travelling inside or outside of their country. Each government needs to monitor every citizen and the COVID-19 test becomes an essential evidence for people who are travelling. These rapid changes are having a serious economic effect on governments and healthcare institutions. This drives the necessity of new digital technologies. It is

---

<sup>1</sup> Ioannis Karamitsos, Rochester Institute of Technology- Dubai Campus, Silicon Oasis, 341055 Dubai, United Arab Emirates; E-mail: ixkcad1@rit.edu.

anticipated that new technologies such as blockchain will increase effectiveness, transparency, and trust of data sharing and will shake the policies, health protocols and procedures of the government.

The digital transformation and digital technologies produced a paradigm shift in the healthcare industry. The outcome of digitalization produces benefits for the healthcare systems and electronic medical test records improve the access to the health records both for patients and health practitioners. On account of the huge demand for the COVID-19 tests and the prevention of disease dispersion there is a clear demand for a new blockchain decentralized global healthcare approach. A COVID-19 test result is the health digital certificate of each person, and it is available to all responsible parties inside or outside countries. The availability of these health certificates is provided using Blockchain technology decentralized platforms.

Blockchain technology is considered as one of the major breakthroughs since the invention of the market. A forecast done by Gartner revealed that the business value-added of blockchain will reach \$176 billion by 2025 and \$3.1 trillion by 2030 [3]. The blockchain technology allows the exchange of a value without the need of a central authority [4]. A number of studies have found that blockchain is a new disruptive technology within the digital era [5] and avoiding the middleman. This make the technology quite cost effective [6] and minimize the risks and conflicts developing between the actors during the transaction execution. In addition, blockchain keeps a digital copy of the COVID-19 test certificates owners with an encrypted signature (hash key) into a local database for each node.

These digital certificates that are stored on blockchain storage hold plenty of advantages over the “ordinary” digital certificates in the following ways:

- **IMMUTABILITY:** COVID-19 certificates cannot be forged, and their content cannot be altered by external parties. The correctness of the data can be checked by the test certificate owner.
- **DECENTRALIZED:** COVID-19 certificates were approved by different actors who have read and write access to blockchain nodes. In consequence, there is no need for any intermediary and the centralized approach is eliminated.
- **VALIDATION:** COVID-19 test certificates maintain their validity status without any third party’s intervention.

The hash is a secret key associated to the original document that the citizen/user keeps. This means that you can publish the signature of a document without publishing the document itself, thus protecting the privacy of the document. Another method is used to issue COVID-19 certificates directly to blockchain network for validation and verification. One advantage of the direct issuance of COVID-19 certificates is that it avoids the problem of double verification. The main disadvantage of this method is that the scalability problem, thus effecting the low performance and high usage of the blockchain. Therefore, opting for a blockchain platform is an issue and depends on several parameters.

This study provides an exciting opportunity to advance our knowledge of understanding how blockchain and health digital certificates can interwork to allow this technology to be successfully implemented in healthcare institutions.

The paper offers a background information about the importance and the surprise of the blockchain technology followed by the description of the health COVID-19 test certificates and the implementation in smart contracts. The readers will have the opportunity to see how this solution operates in practice and how the health digital COVID-19 tests are fully utilised using Hyperledger and Ethereum platforms. The

selection of the platforms depends on the health protocols that each government has applied.

The remaining of the paper proceeds as follows: Section 2 presents the literature review using bibliometric techniques. Section 3 begins by developing COVID-19 digital test certificates using the Hyperledger platform. Section 4 presents the Ethereum platform, focusing on the smart contract implementation. Section 5 draws the citizen registration process as part of the whole healthcare pipeline. Finally, the conclusion gives a brief summary of the findings and recommendations for future research in this area.

## 2. Related Work

The study uses a systematic review in order to search the literature review and identify gaps by using different keywords: digital certificates, blockchain and COVID-19. We used a bibliometrix tool developed by Aria Massimo [7] to search for conference papers and journal articles using the scopus database. A number of 245 papers were identified for further analysis. After collection, the papers were examined on the basis of a degree of homogeneity of their title and abstract and the papers were reduced to 59 papers. The results in this literature review indicate that blockchain technology can help the pandemic minimize the infections rates and the economy optimistically. There is an unambiguous relationship between blockchain, COVID-19 and healthcare transmission stages. It is interesting to note that the distribution and access of electronic medical records (EMR) are becoming essential after the COVID-19 pandemic. Yang et al. [8] concludes that healthcare management information is far most cost effective, and therefore better value-added to the healthcare processes. Previous research has indicated that smart contracts have a positive impact on the automatic execution of transactions and information exchange (Karamitsos et al., [9]). In addition, there was a significant positive association between smart contracts and COVID-19.

Christ et al., [10] stated that current medical and healthcare providers are insufficient. Health problems seem to be becoming more complex, necessitating a sufficient healthcare system. Instead of availability and traceability of data, IT in health industry is improving electronic medical records, biomedical databases, and public health. Due to privacy concerns, most healthcare institutions keep archives of patient records that are unavailable to other institutions despite the fact that studies have shown that EMR exchange can help resolve healthcare issues. Health care institutions keep these essential data private [10]. Implementing blockchain in health industry benefits in multiple ways to the patients as well as the institution i.e., - accessing their health records, granting permissions to institution to get access to the patient's data, enhanced pharmaceutical supply chain management processes, and patients' data can be analyzed for medical innovations and research results. However, there is a gap for the verification and ownership of the electronic medical records [11].

Ting et al. [12] investigated the differential impact of digital technologies on public health administrations. By drawing on the concept of blockchain (Ting et al. [12]) have been able to show a collaborating system between different stakeholders such as patients, pharmacists, and hospitals for accurate tracking.

One study of Chamola et al. [13] examined the usage of disruptive technologies such as artificial intelligence (AI) and blockchain in order to minimize the impact of the pandemic. The development of mobile applications has been widely investigated for storing health medical data in immutable records. In their research study, Chamola et al.

[13] identified two mobile applications which include MiPasa and CIVITAS. MiPasa application was designed in Hyperledger fabric platform in order to gain insights into the patient personal information. CIVITAS application used for filling the healthcare records capturing citizen's ID personal information. The healthcare records were transferred and loaded into the blockchain storage providing ownership to citizens [13].

Mashamba et al. [14] provided an in-depth analysis of the tracking system showing its relevance to infectious diseases using both blockchain and artificial intelligence. The tracking system development is based on Geographic Information System (GIS) installed in smartphones enabling tracking between positive tested.

Vaishya et al. [15] draw our attention to disruptive technologies for combating the COVID-19 pandemic disease. In their review of blockchain processes presented the benefits of traceability and real-time information [15].

In summary, these studies show that blockchain is a disruptive technology providing secure verification regarding the citizen information [12-14]. The citizen information is stored at every blockchain node in a decentralized topology providing tamper-proof and immutability. Further investigation is required to determine the blockchain limitations such as performance, legal regulation and limited technology awareness.

### **3. Hyperledger Blockchain Platform**

Hyperledger [16] is an open source software under the Linux Foundation. The use of Hyperledger platform is a well-established approach in blockchain and consists of infrastructure, frameworks and tools. Hyperledger fabric framework is selected as the primary platform for the COVID-19 digital test certificate. The Hyperledger, unlike the bitcoin and Ethereum did not require any mining process for the validation of their transactions. The benefits of the selection of the Hyperledger fabric platform are presented as follows:

- Hyperledger Fabric issues transactions with derived certificates that are unlinkable to the owning participant.
- Relies on a smart contract system (Chaincode), runs in Docker containers by any peer of the network.
- All participants must register proof of identity to membership services in order to gain access to the system.
- Each transaction content was secured using an encryption mechanism to ensure that the content is available only to the intended participants.
- Hyperledger fabric contains only secure, private, and confidential transaction. Only peer members can modify the content of Hyperledger fabric.
- Transactions are executed without a mining validation.
- The events are structured as transactions and are exchanged between the different participants.

The Hyperledger fabric application has three main components. First, the model file which contains the definition of each class of assets, transactions, participants and events. Table 1 shows an overview of roles and responsibilities for each participant.

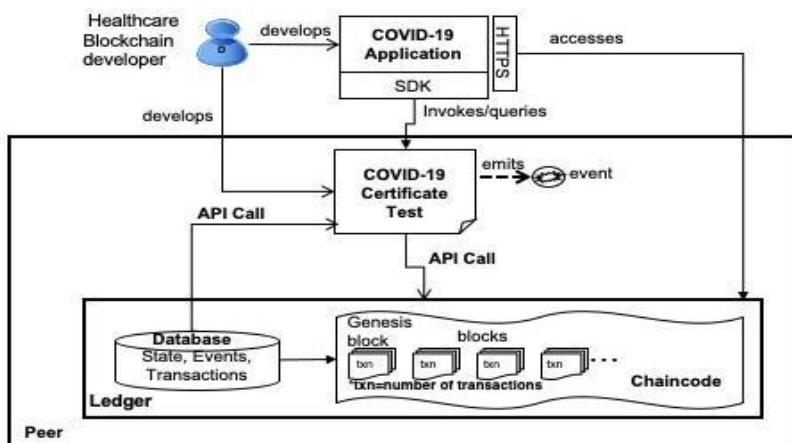
**Table 1.** Hyperledger Fabric Participants/Actors

Hyperledger Participants	Roles /Responsibilities
Healthcare Institutions (Blockchain Developer/Architect)	Application Design & Development/Big data Storage for Covid19 tests/ Issuance of COVID-19 test certificate template
Healthcare Regulator	Monitoring the health protocols
Government Institutions	Issuance of COVID19 test certificate at country level
Citizens/Users	Provide personal information and ownership of Covid-19 Test Certificate

Second the script file: It is also called the transaction processor function file and it houses the JavaScript logic to execute the transaction explained in the model file. The two transactions performed in this paper are the publication of COVID-19 certificates by the healthcare institutions and the issuing of a certificate by the citizen to a beneficiary. Third, the ACL file: This contains the data within a business network which defines an asset or groups of assets as well as the participants that can conduct operations that affect those assets. The asset for this research is the certificate that contains the citizen details and the test results of COVID-19.

### 3.1. Hyperledger Implementation

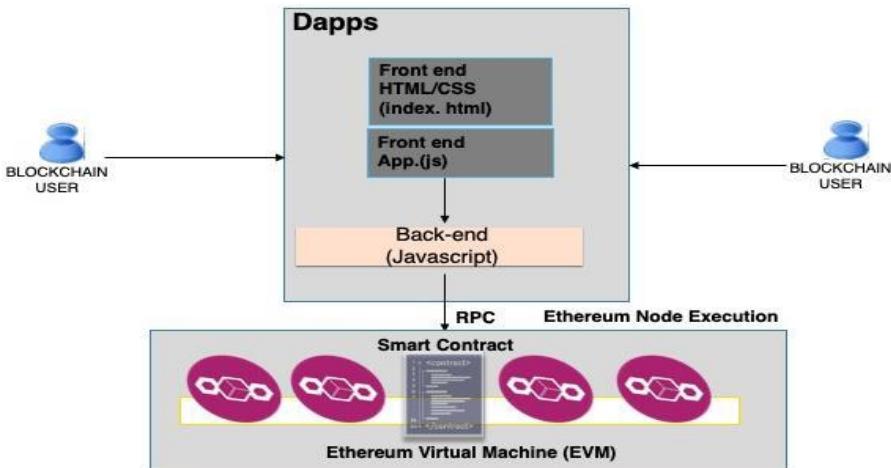
The chaincode [16] component plays the role of the smart contract within the Hyperledger platform. The healthcare developers shall design and develop an application and the smart contract (chaincode) as presented in the following Figure 1. Chaincode [16] is distributed inside the blockchain network and control the ledger's status. The application handles user interface and submits chaincode named transactions to the network.

**Figure 1.** Hyperledger Implementation

Therefore, the blockchain network emits transaction block events which enable applications to integrate with other systems.

#### 4. Ethereum Platform

Ethereum [17] is a technology for smart contract development. Smart contracts are complex set of software code with components designed to automate execution on the blockchain and are capable of imposing rules, consequences and computation over any transaction that happens in the blockchain. The study of the smart contract was first carried out by Nick Szabo (1997) [18]. Smart contracts may take any type of data as input, as the COVID-19 test certificate in our study. In addition, smart contracts based on the protocols defined in the smart contract perform input computation and can implement decisions based on the prevailing output production conditions. Functions and state machine are the main components of the transactions. Smart contracts are implemented by Turing complete programming languages which can translate the contract into bytecode and execute it using Ethereum Virtual Machine (EVM) [9] as depicted in Figure 2.



**Figure 2.** Ethereum platform Implementation

The Ethereum architecture and implementation are completely independent of the bitcoin cryptocurrency. Solidity (Java-like) [19] and Serpent (Python-like) [20] are currently available programming languages which support the Ethereum platform. Smart contracts and decentralized applications (Dapps) are written with a high-level programming language. The developer will create the format of transactions, the functions of events and state transitions, and the rules of ownership. Ethereum Virtual Machine (EVM) is a virtual machine in which the software code is executed. [9].

##### 4.1. Ethereum Implementation

The design and development of the smart contract was carried out over the Ethereum platform for the implementation of COVID-19 digital certificates. COVID-19 smart contract architecture is comprised of the following steps:

First, the Ethereum node setup with actors and roles is required for every citizen/user and healthcare institutions; second, the business services/functions are

specified and finally, the collaboration among the users and services/functions is implemented. The above steps are discussed in detail in the following sections.

#### 4.1.1. Ethereum Node Setup

The setup of the Ethereum node was based on actors, business services and Ethereum processes.

##### 4.1.1.1. Actors and Roles

Actors and roles were identified using the COVID-19 smart contract developed in Ethereum node as follows:

- Smart contract owner is typically healthcare government or private institutions that are responsible for creating the smart contract and have a digital certificate from COVID-19 test.
- Citizens/users are responsible for building their Ethereum wallets in the public/private blockchain and have access to P2P nodes.

##### 4.1.1.2. Business Services and Functions

The business services and functions which the smart contract needs are as follows:

- Registration of citizens/users
- Design and develop Smart Contract
- Create functions/events
- Send messages/events
- Use cryptocurrency (Ether or BTC) for gas mining fee

##### 4.1.1.3. Ethereum Processes

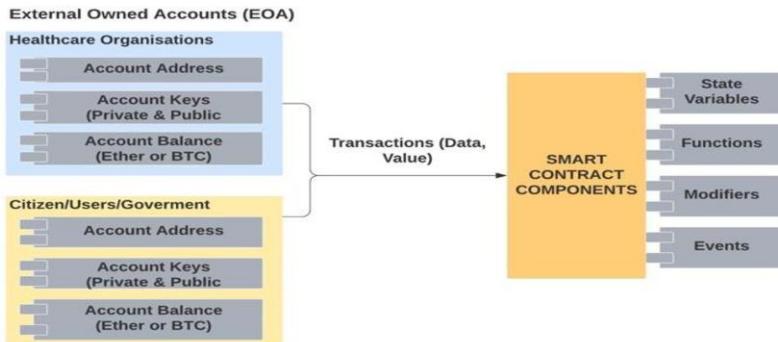
We have four processes for the Ethereum platform, which are:

- Block validation process
- Network discovery process: for a new node to join the P2P blockchain network
- The design of transactions enabling citizens and healthcare organisations to make transactions and enabling COVID-19 smart contract to generate events and messages.
- The transaction validation process (COVID-19 test certificate) into the block using mining approach.

#### 4.1.2. Smart Contract

For the Ethereum framework, transactions were executed using a group of functions and state variables written in software programming languages. Each transaction contains input parameters which involve a smart contract function. Depending on the logic implementation the state variables changed status during the execution of a function.

Using the programming languages Solidity [19] or Serpent [20], both the functions and the state variables are translated into program code. The contract code is then submitted to the blockchain network until the compiler runs without any errors. The blockchain network will assign a unique address for each contract that is submitted. Any citizen/user or healthcare organization can submit events to contracts other than this. The event consists of the sender's address, the recipient's address, the transfer value, and a data field containing the receiver contract input data. There is a distinction between transaction and event where the smart contract generates the event while the transaction is generated by the External Owned Account (EOA) [9], as shown in Figure 3.

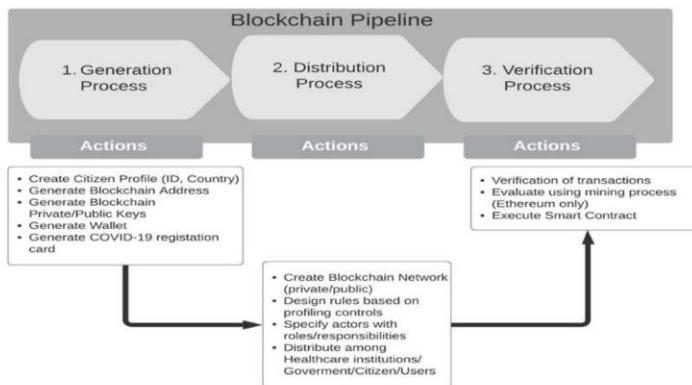


**Figure 3.** Smart Contract

Healthcare organisations and citizens/users/government have been identified as external owned accounts (EOAs) for the smart contract. The external owned accounts were managed using security mechanism (private and public keys) and smart contract transactions were transferred data and value. Smart contract accounts (SCAs) management were performed using their software code. The smart contract messaging system executed in timely manner, provided read and write access to the internal database and communicated with other smart contracts. Miners validate the blocks and transactions. The transaction is enclosed in a block, and for this block, miners will validate the transactions using a consensus mechanism, for example proof of work (PoW). After the transaction is authenticated into the chain, miners are given a sum as a reward.

## 5. COVID-19 Test Certificates Pipeline

In any blockchain project, teams of developers and healthcare practitioners work to build state-of-art models using automatic driven processes. COVID-19 test certificates pipeline process is a convenient and frequently used tool in the study of COVID-19 use cases. The process is script-driven, and every step is automatic including generation, distribution, and verification stages. The COVID-19 test certificates pipeline process is composed of three stages as depicted in Figure 4.



**Figure 4.** Blockchain Pipeline

From the abovementioned Figure 4 we select for further analysis the generation of the citizen registration process. This process is the most important step for the COVID-19 test certificates because we have the involvement of the healthcare institutions and the citizen.

### 5.1. Use case: Citizen Registration Process

This process can be used for both blockchain platforms. The two main actors are the citizen and the healthcare blockchain administration. Figure 5 demonstrates how blockchain can enhance mobility between healthcare institutions, government and citizens.

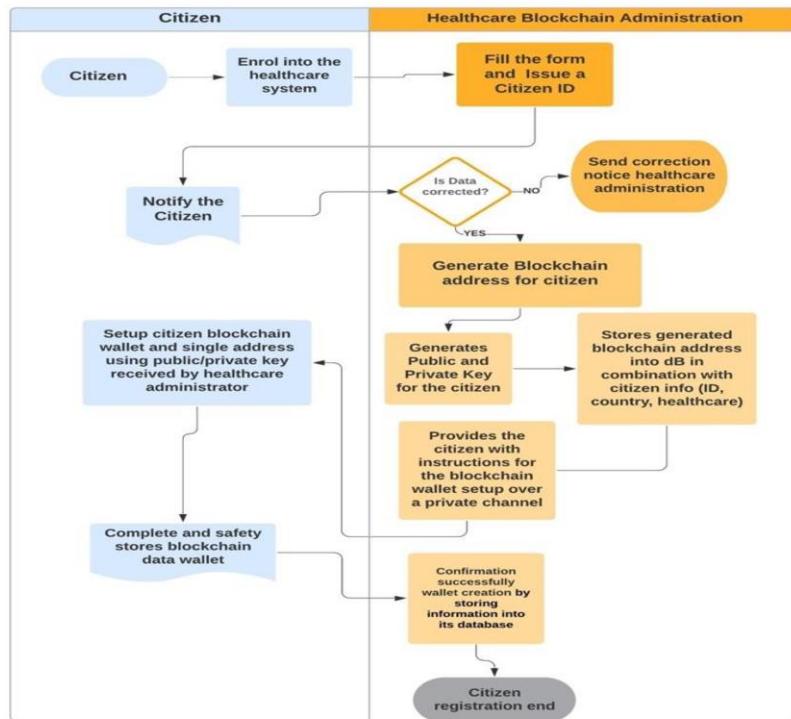


Figure 5. Citizen Registration for COVID-19 Digital Certificate

The citizen registers in the country healthcare platform introducing his personal data and the healthcare national number. Upon receiving an application submission context from a citizen, the healthcare national system produces a unique citizen ID. The system notifies the citizen by SMS or email. The citizen checks the correctness of the data file. If the data is correct, then the healthcare system generates a blockchain address with public and private keys. Then the citizen creates a blockchain wallet using public/private keys from the system. This information is stored safely in the citizen's wallet. A blockchain synchronization is performed between the citizen and the healthcare system.

## 6. Conclusion and Future Work

The present study was designed to determine the effect of the blockchain platforms in healthcare sector. This research was undertaken to design a smart contract and implement a use case of COVID-19 test certificates. Healthcare organisations should assess whether and when blockchain can be used as a solution in their ecosystem.

The present work makes several noteworthy contributions to COVID-19 pandemic using smart contract and blockchain platforms. A variety of advantages are presented as follows:

- Equitable access: COVID-19 test results can be accessed by different actors. In the healthcare ecosystem multiple actors such as healthcare organisations, regulator and citizen/user involve the management of the COVID-19 certificate. They have access to read/write or modify any information depending on the access rights provided by the blockchain owner, which is in our case healthcare organisations that are responsible for the ownership of the COVID-19 test results.
- Low risk fraud: Distributed ledgers were chosen for the storage of COVID-19 test certificates and copies of transactions are distributed among the nodes in the network. This gives the blockchain an attribute of making it almost impossible to alter the data by external parties. Therefore, the COVID-19 test certificates maintain the immutability property of the blockchain technology.
- Privacy: The smart contract for COVID-19 offers a secured privacy to citizens/users. The COVID-19 test results are under citizens' control and not giving the privilege to a certain institution to use this data which may have value to another external party.
- Security: Blockchain platforms and decentralized ledgers use an encrypted database providing more resilience and less exposure to malicious attackers.
- Efficiency: COVID-19 certificates were generated a common version for all the transactions across the network, blockchain platforms provide instant visibility.

A smart contract structure for Hyperledger and Ethereum platforms is presented in detail. Further studies, which take these blockchain platforms into account, will need to be undertaken before the application of healthcare protocols applied in different countries. Also, a common healthcare COVID-19 test template is required to design and develop a smart contract.

## References

- [1] Mayo Clinic, USA. Retrieved on 12 July 2020 from: <https://www.mayoclinic.org/diseases-conditions/coronavirus/symptoms-causes/syc-20479963>
- [2] World Health Organization Coronavirus 2019 (COVID-19) situation report-51. World Health Organization (2020). Retrieved on 22 July 2020 from: [https://www.who.int/docs/default-source/wha-70-and-phe/20200721-covid-19-sitrep-183.pdf?sfvrsn=b3869b3\\_2](https://www.who.int/docs/default-source/wha-70-and-phe/20200721-covid-19-sitrep-183.pdf?sfvrsn=b3869b3_2)
- [3] Gartner Research (2017). Forecast: Blockchain Business Value. *Worldwide, 2017-2030*.
- [4] Niranjanamurthy, M., Nithya, B. N., & Jagannatha, S., (2018). Analysis of Blockchain technology: pros, cons and SWOT. *Cluster Computing*
- [5] H. Kakavand, H., N. Kost De Serves, N., & Chilton, B.,(2017). The Blockchain Revolution: An Analysis of Regulation and Technology Related to Distributed Ledger Technologies. *SSRN Electron. J.*

- [6] Kosba, A., Miller, A., Shi, E., Wen, Z., & Papamanthou, C. (2016). Hawk: The Blockchain Model of Cryptography and Privacy-Preserving Smart Contracts," in *Proceedings-2016, IEEE Symposium on Security and Privacy*, SP 2016.
- [7] Aria, M. & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis, *Journal of Informetrics*, 11(4), pp 959-975, Elsevier, <https://doi.org/10.1016/j.joi.2017.08.007>
- [8] Yang, H., & Yang, B. (2017). A Blockchain-based Approach to the Secure Sharing of Healthcare Data. *Norwegian Information Security Conference*.
- [9] Karamitsos, I., Papadaki, M., & Barghuthi, N. (2018). Design of the Blockchain Smart Contract: A Use Case for Real Estate. *Journal of Information Security*, vol.9, pp.177-190.
- [10] Christ, M.J., Nikolaus Permana Tri, R., Chandra, W., & Gunawan,W.(2019). Exploring Blockchain in Healthcare Industry. In *2019 International Conference on ICT for Smart Society (ICISS)*, pp. 1-4, <https://doi.org/10.1109/ICISS48059.2019.8969791>.
- [11] Al-Jaroodi, J., & Mohamed, N. (2019). Blockchain in Industries: A Survey. *IEEE Access*, 7, pp. 36500–36515
- [12] Ting, D. S. W., Carin, L., Dzau, V., & Wong, T. Y. (2020). Digital technology and COVID-19. In *Nature Medicine*. <https://doi.org/10.1038/s41591-020-0824-5>
- [13] Chamola, V., Hassija, V., Gupta, V., & Guizani, M. (2020). A Comprehensive Review of the COVID-19 Pandemic and the Role of IoT, Drones, AI, Blockchain, and 5G in Managing its Impact. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2020.2992341>
- [14] Mashamba-Thompson, T. P., & Crayton, E. D. (2020). Blockchain and artificial intelligence technology for novel coronavirus disease-19 self-testing. In *Diagnostics*. <https://doi.org/10.3390/diagnostics10040198>
- [15] Vaishya, R., Haleem, A., Vaish, A., & Javaid, M. (2020). Emerging Technologies to Combat the COVID-19 Pandemic. In *Journal of Clinical and Experimental Hepatology*. <https://doi.org/10.1016/j.jceh.2020.04.019>
- [16] Linux Foundation, About Hyperledger Fabric. Retrieved on 22 July 2020 from: <https://www.hyperledger.org/projects/fabric>
- [17] V. Buterin, "On public and private blockchain.", Ethereum blog, crypto, (2015). Retrieved on 22 July 2020 from: <https://blog.ethereum.org/2015/08/07/on-public-and-privateblockchains/>.
- [18] N. Szabo, "Formalizing and securing relationships on public networks," First Monday 2(9),1997. Retrieved on 22 July 2020 from: <https://doi.org/10.5210/fm.v2i9>.
- [19] Solidity documentation. Retrieved on 15 July 2020 from: <https://solidity.readthedocs.io>
- [20] Serpent documentation. Retrieved on 15 July 2020 from: <https://ethereumclassic.org/blog/2017-02-10-serpent>

# Assessment of National Cybersecurity Capacity for Countries in a Transitional Phase: The Spring Land Case Study

Mohamed Altaher Ben Naseir <sup>a</sup>, Huseyin Dogan <sup>a<sup>1</sup></sup>, and Edward Apeh <sup>a</sup>

<sup>a</sup> Bournemouth University, Fern Barrow, Poole, Dorset, BH12 5BB, United Kingdom

**Abstract.** Cybersecurity capacity building has emerged as a notable matter for numerous jurisdictions. Cyber-related threats are posing an ever-greater risk to national security for all countries, irrespective of whether they are developed or in the midst of transitioning. This paper presents the results of two qualitative studies using the Cybersecurity Capacity Maturity Model (CCMM) for nations: (1) Interactive Management (IM) and (2) focus groups to analyse the current state of Spring Land's cybersecurity capacity. A total of 26 participants from government agencies and five national experts from the Spring Land National Cybersecurity Authority (NCSA) contributed to this study. The results show that Spring Land has many issues such as lack of cybersecurity culture and collaborative road-map across government sectors which results in instability within the country. The assessments feed into the requirement analysis of the National Cybersecurity Capacity Building Framework that can be utilised to organise and test the cybersecurity for nations.

**Keywords:** Cybersecurity capacity, Cybersecurity Maturity Models and Interactive Management.

## 1. Introduction

Over numerous decades, there have been several notable security failings that have defined the global security environment and resulted in governments being unable to preserve domestic security [1]. Maintaining national security (including cybersecurity) is the main responsibility of national governments and failing to do so contributes to the instability of a country [2]. Countries in a transition phase is typically characterised by civil war; political and economic upheaval; the absence of law [3, 4]. Transition phase refers to the intermediate phase in which a previous regime is replaced by a modern alternative [5]. There are a number of factors affecting the success or failure of transition stage. These factors include: the type of regime prior to the transition stage,

---

<sup>1</sup> Corresponding Author, Huseyin Dogan, Bournemouth University, Fern Barrow, Poole, Dorset, BH12 5BB, United Kingdom; Email: hdogan@bournemouth.ac.uk

the characteristics of the new leader of the transitional government and the influence of information and communication technologies (ICT) [6].

Many of these countries depend on cyberspace to provide daily services for their citizens using information and communication technologies (ICT). The growth of ICT technologies and applications provides an important vehicle for communication and interaction and has increasingly become common in low-income countries and countries that are in a transitional stage [7]. Access to a range of ICTs brings new opportunities for information exchange and communication but it also can also be seen as a technological and generational challenge to the hierarchical social order of many countries in the Middle East and North Africa (MENA) region. The current experience of democracy movements (the Arab Spring) in a number of MENA countries demonstrates how during times of public protest and turbulence, ICTs can be significant forces for organisation and mobilisation [7].

This paper aims to evaluate the capacity of cybersecurity in countries progressing through a phase of transition by taking Spring Land as an exemplar case study. The name ‘Spring Land’ has been selected to disguise the real name of the country in which the case study has been undertaken.

The assessment was undertaken by applying the Cybersecurity Capacity Maturity Model (CCMM) - V1.2, utilising the Interactive Management (IM) approach and focus group discussion method. The CCMM model was designed by the Global Cybersecurity Capacity Centre at the University of Oxford [8]. The CCMM model was nominated because it successfully demonstrates the effect that a Cybersecurity Capacity Building (CCB) approach can achieve at the worldwide level, including all aspects of cybersecurity to ensure that the platform remains resilient. This assessment has provided a great opportunity to illustrate the fact that in the current hyper connected world, states in a transitional phase are not operating in isolation and their failure in certain critical areas such as cyberspace is likely to have a ripple effect by destabilising stable states. Moreover, it will feed into the requirement analysis of the National Cybersecurity Capacity Building Framework and the possibility of organising and testing cybersecurity in these countries.

The paper is structured as follows: Section 2 discusses the related empirical research; Section 3 provides an overview on the CCMM; Section 4 presents the selected methodology and the problem of space contextualisation through the IM. The cybersecurity posture of Spring Land through the focus group discussion is presented in Section 5. Finally, Section 6 provides a conclusion and recommendations for future research in this area.

## 2. Cybersecurity Capacity Building (CCB)

Cybersecurity Capacity Building (CCB) is one of the greatest challenges that countries face, particularly countries in a transitional stage. These challenges are range from human resource development, institutional reform, organisational adaption, and the support provided to increase their potential to not only make use of the Internet but also realise its full potential [9, 10]. The majority of problems relate to the lack of cybersecurity culture and an inability to understand the threat posed as well as the probable consequences [11].

Furthermore, various other issues can affect decisions when building a secure cyberspace. For instance, many countries lack a legislative framework, the resources required to build what they need and secure capacities in cyberspace. Also, awareness of and education about the threats and risks associated with cyberspace are common issues in these countries. Without awareness and education, attempts to secure a system are rendered inefficient, if not useless [9, 12]. Another problem is linked to the dearth of skills among Internet users to protect themselves against rapidly emerging cyber-threats. In many developing countries and countries in a transitional period, Internet users are inexperienced and are not technically savvy.

The term ‘capacity building’ refers to the process of addressing an identified issue with poor governance by ensuring a suitable capability so that core functions are delivered [11]. Therefore, capacity building entails developing organisational structures (i.e. methods of management at the organisational level), human capital (i.e. addressing skills shortages and enhancing knowledge), and the frameworks that underpin legal and institutional arrangements (i.e. strategies and legislation).

However, various frameworks and guidelines have been devised by academic researchers as well as organisations operating in the country or worldwide. From these, it is apparent that five pillars support cybersecurity capacity: human, organisational, infrastructure, technology, law and regulation [13]. Such frameworks are primarily concerned with the risks to cybersecurity and the steps that can be taken to protect against them at the international level and especially in advanced economies. In addition, it is apparent from the empirical literature that there is a paucity of research focusing on emerging market countries owing to their relative shortage of human capital as well as technical capacity [14]. National governments and international organisations have recognised the threat posed by such risks but efforts to implement effective defences have not been coordinated and this disjointed approach has resulted in certain countries being much better prepared than others [15]. According to Muller [9], current efforts to address CCB have not taken a global perspective or have advocated CCB but failed to suggest how it should be implemented.

### **3. Cybersecurity Capacity Maturity Model for Nations (CCMM)**

The CCMM was developed by the Global Cybersecurity Capacity Centre at the University of Oxford through collaboration with international stakeholders including the Organization of American States (OAS), the World Bank, the Commonwealth Telecommunications Organisation (CTO) and the International Telecommunication Union (ITU) [8]. The model offers a comprehensive analysis of cybersecurity capacity through five dimensions. These dimensions are cybersecurity strategy, Cybersecurity awareness, Cybersecurity education, training and skills, Cybersecurity legal framework and the Standards, Organisations and Technologies.

Each dimension has multiple factors which define what it means to possess a cybersecurity capacity. For each factor, there are five stages of maturity. The Start-up indicator describes a non-existent or inadequate level of capacity; the Formative level indicates that some features are formulated but poorly defined; the Established pointer shows that an element of the sub-factors are in place and defined; in the Strategic indicators level the selections of which parts of indicators are vital or less important have been made for particular institutions/nations based on certain conditions; the Dynamic indicator level is the highest level and indicates that there are clear mechanisms in

place to modify the strategy subject to the prevailing circumstance. The results of the maturity levels are graphically represented using a radar chart [8].

## 4. Methodology

The assessment of the national cybersecurity capacity of Spring Land utilises two qualitative approaches: Interactive Management (IM) and focus groups discussions using the CCMM for Nation states. In this study, in order to gain a more thorough understanding of the Spring Land cybersecurity posture in which the model will be applied, the authors worked alongside the Spring Land National Cybersecurity Authority (NCSA). The NCSA leads the national cybersecurity programme in Spring Land to achieve resilience in cyberspace [16]. The following sections provide more details about the methodology used in this paper and the participant's profile.

### 4.1. Interactive management

The IM approach relates to complicated scenarios that demand collaboration among numerous knowledgeable individuals to address the matter and suggest a plan of action based on mutual agreement instead of a majority vote[17, 18]. There are three phases in IM, the first of which is the planning phase where the scenario and scope are specified. This involves creating a formal scope and context statement, defining the state of assessment, and verifying the identities of the related actors. During the workshop phase, the participants develop a shared understanding of events [17].

There are three procedures involved in IM workshops: idea writing (IW); nominal group technique (NGT); and interpretive structural modelling (ISM) [17, 18]. IW involves the participants being presented with a question so that they can develop their thoughts in writing and only then share their ideas. During the NGT, those participating assess the matter from a holistic perspective based on what occurred during the IW process. A ranking of the various ideas is compiled on the basis of their importance. The idea statements are then used as the basis for developing objectives and an Interpretive Structural Model (ISM) so that the way in which the factors associated with the problem relate to each other is recognised. In the follow-up process, the objectives and outcomes previously arrived at are acted upon to help bring about a viable solution. The authors had selected this method because IM sessions are conducted as part of an integrated approach for dealing with the situation, and each session builds on what came before and lays the foundation for what will come after [17].

In this study, a one-day workshop was hosted by NCSA for a total of 26 participants representing various stakeholders. The information details of the participants involved in the workshop are described in the participant's profile section. The results of this approach were published at the World Conference on Information Systems and Technologies (2019) [10].

### 4.2. Focus group

Focus group discussions aim to explore a various opinions that people have regarding particular matters and emphasising the different thoughts that groups of people have [19]. Conducting a focus group entails people collaborating about a particular subject

matter to enable the collection of relevant data [20]. Focus groups give people the opportunity to interact with each other in a way that yields useful information and a range of opinions. The decision was taken to hold focus groups because the authors believed it would generate richer data than would otherwise be possible if selecting alternative methods. [21]. In this study, five experts from the NCSA were interviewed in one session hosted in the capital city of Spring Land.

#### 4.3. Participant's profile

Two workshops hosted by the NCSA were conducted with a national expert from Spring Land. The IM approach was conducted with a total of 26 participants from different stakeholders, 25 males and 1 female only due to lack of gender diversity involved in cybersecurity roles. The ages of those participating were within the range of 25-55 years and they had been selected because of the contribution they make to decision-making processes. They were drawn from various areas of expertise including banking, management, defence, security, oil production, immigration, digital crime and the intelligence service.

The focus group discussion was conducted with five experts (lead practitioners) from the NCSA. The participants (Ps) in this session were chosen based on their roles within the NCSA. The participants comprised senior management of the NCSA in Spring Land, a director of the NCSA (P1), a deputy director of the NCSA (P2), the head of the national cybersecurity incidents response team (CERT) (P3), the head of awareness and general relations (P4), and the head of the internal audit office (P5). For the purposes of confidentiality, the names of the participants were not disclosed.

## 5. Results

### 5.1 Problem space contextualisation through Interactive Management

#### 5.1.1. Ideas writing (IW) results

An IW was employed to identify matters associated with a particular trigger question, thereby enabling those participating to share opinions and brainstorm in a group setting. Those participating were assigned to one of three groups where they discussed the question and offered opinions relating to the state of Spring Land's cyber security. The selected trigger question sought to identify the cybersecurity capacity issues faced by Spring Land. The trigger question employed was: *What are the current issues of cybersecurity capacity in Spring Land?*

Once the session had concluded, each of the statements that had been made were assigned a number and categorised on the basis of the CCMM dimensions. The ideas generated by the groups in response to the question are summarised and Table 1 presents examples of the challenges of cybersecurity capacity in Spring Land.

**Table 1.** Examples of national cybersecurity capacity challenges of countries in transitional stage vs CCMM Dimensions[10]

D1 - Cybersecurity policy and strategy	D2 - Cyber culture and society
D1.1. Absence of a national cybersecurity strategy.	D2.1. Lack of a cybersecurity culture and the absence of an understanding of cyber-risk and its consequences in the public and private sectors as
D1.2. Unavailability of a national risk management plan and threat of cyberspace has not been identi-	

fied at the national or sector-specific level.

D1.3. Deficiency of a national roadmap for a cyber defence strategy.

well as among decision-makers.

D2.2. Lack of awareness-raising programmes at the governmental level.

D2.3. Citizens' confidence in the use of e-government services is weak.

### 5.1.2. Nominal group technique (NGT) results.

The purpose of using the NGT was to produce, simplify and amend ratings for a series of objectives. Those participating chose what they believed to be the three main objectives for the various dimensions (where 1 indicates the lowest importance and 3 indicates the greatest importance). 19 of the participants cast their vote but 7 did not because of external commitments or other reasons. Figure 1 illustrates the objectives of greatest importance for the various dimensions as well as indicating how they interrelate.

### 5.1.3. Interpretive structural modelling (ISM) for countries in transitional stage.

The ISM approach enabled those participating to analyse how the elements resulting from the NGT process are inter-related, providing the means to address the associated complexities [22, 23]. So as to ensure that the ISM is clear, the objectives of the NGT stage were categorised on the basis of their similarities to help identify the most notable objectives of the respective dimensions. Figure 1 illustrates the ISM resulting from the objective statements and how they interact on the basis of the CCMM's dimensions.

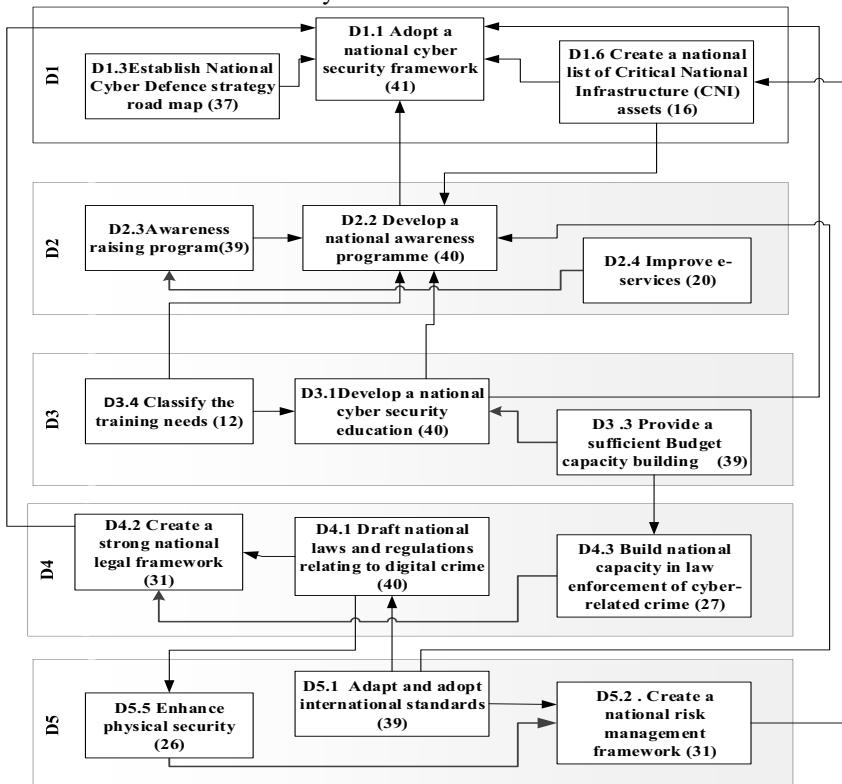


Figure 1. Interpretive structural modelling for countries in a transitional phase

It is apparent from Figure 1 that there is a need for a national blueprint owing to the fact that the existing state cyberspace interactions make clear the insufficient capacity of national cybersecurity. It can also be seen that the group believed providing a comprehensive programme to boost awareness throughout the country would help to enhance the capacity of cybersecurity at the national level. It was the group's belief that establishing a national strategy would help to initiate a process that would result in a national legal framework to enhance the sharing of information, the disclosure of susceptibilities and reporting among public sector bodies. Furthermore, it was recommended by the group that international standards (e.g., ISO27000) be adopted by government bodies so as to bolster efforts to ensure effective technical controls. In addition, it was claimed that the capacity to tackle threats (both internal and external) would benefit from improvements to physical security.

## 5.2 Spring Land's cybersecurity posture through focus group discussion

This section discusses the results of applying the CCMM model to assess Spring Land's cybersecurity posture using a focus group discussion approach. Five experts (lead practitioners) from the NCSA participated in this discussion, as mentioned in the participant's profile section. The cybersecurity posture of a nation state of the five dimensions of the CCMM model is presented in Figure 2 which shows the overall capacity level results using the radar chart.

### 5.2.1. Cybersecurity policy and strategy indicators (D1).

This dimension explores the capacity of the government to design, create, organise and implement the cybersecurity strategy. Through the discussion, this dimension was classified to be at the start-up stage in Spring Land because no national cybersecurity strategy currently exists. Therefore, NCSA has been assigned to be in charge of the cybersecurity programme.

*"NCSA leads the security of information as there is no single body or group related to cybersecurity in Spring Land. In general, we can say that we are at a strategic level with a total lack of financial support because of the political situation" (P1).*

NCSA created a national computer emergency response team (Spring Land - CERT) which is working only at the level of NCSA departments due to a lack of co-operation, trust, national strategy and poor awareness at the state level. *"In general, we have national accreditation to represent Spring Land in the world but there is no national plan and poor communication channels due to the fear of dealing with one another. In contrast, there is good cooperation at an international level as Spring Land is a member of different international organisations such as ITU" (P3).*

Regarding the critical national infrastructure (CNI) protection, most of the Spring Land critical systems have been destroyed and the government has not issued a list of CNI.

*"In general, physical security has a negative impact on CNI and there are no clear processes to reveal who is in charge of protecting all sectors, except the telecommunications sector" (P5).*

Furthermore, the difficult economic situation and scarcity of means in the country prevents the NCSA from raising awareness and improving national infrastructure protection.

### 5.2.2. Cyber culture and society indicators (D2)

Cyber culture and society at both the individual and government level are at the start-up stage in Spring Land. In the meantime, the NCSA has tried to improve the knowledge base and raise awareness of cybersecurity issues through campaigns and programmes targeted at children, their parents and university students.

*“NCSA has conducted awareness activities for the government sector. As a result of these activities, NCSA reported a lack of awareness programmes in all government sectors and society” (P4).*

Additionally, the NCSA has a plan to change the cybersecurity mind-set and raise awareness of the Spring Land public and national sector regarding spam, scams, phishing, information security, wireless network security and cloud computing security. The NCSA team pointed out that a lack of skilled people and cybersecurity awareness leads to more cybersecurity threats and increased cyber vulnerabilities. Despite the fact that some e-government services in Spring Land have been developed and implemented, a lack of trust and confidence in online security prevails due to there being no online protection across the majority of the government sector.

*“Spring Land has been considered as a target for e-hunting; these are hackers from inside and outside the state. These hackers are creating fake social media pages to commit frauds. There are no public key infrastructures or digital certificates to protect it” (P5).*

Due to political issues and the absence of a legislative body in Spring Land, the maturity of privacy online is considered to be at a start-up stage because no official initiatives have been issued. The exception is a certain unofficial initiative to issue laws for electronic transactions.

### 5.2.3. Cybersecurity education, training and skills indicators (D3)

Throughout the discussion, it has been noted that cybersecurity education, training, and skills capacity in Spring Land is at a start-up level. There are no plans at the national level to increase the efficiency of education in the field of cybersecurity.

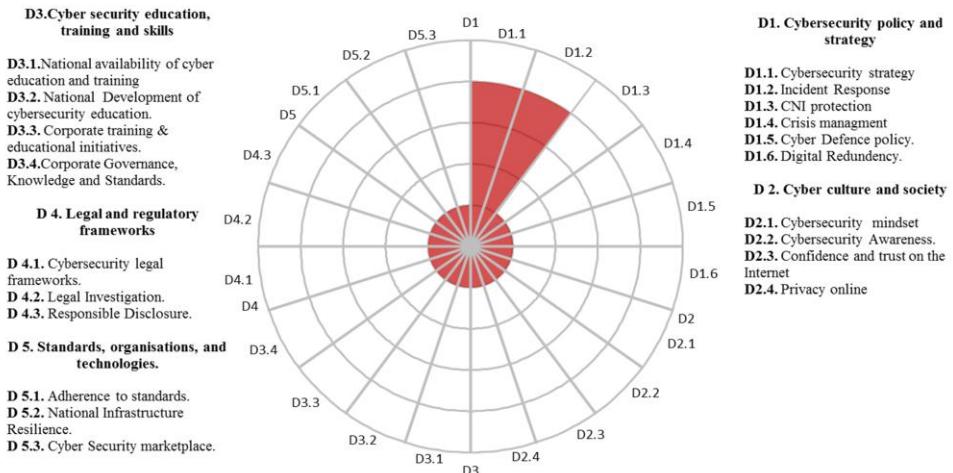
*“There are no plans at the state level to define the required educational curricula in cybersecurity” (P4).*

Additionally, there are no current or future financial allocations, co-ordination or training plans between universities and the private sector regarding cybersecurity training at the state level due to a lack of interest.

### 5.2.4. Legal and regulatory framework indicators (D4)

This dimension looks at the government’s capacity to design and develop national legislation and accompanying by-laws that directly and indirectly relate to cybersecurity. In Spring Land, the level of maturity for this dimension is considered to be at the start-up phase. There is no cyber- or ICT security-related legislation or regulations except for some initiatives by the e-Commerce Chamber of the Ministry of Economy. These initiatives face many problems, but the crucial problem is jurisdictional fragmentation due to political instability. Moreover, there is a digital crime unit in the Ministry of the Interior that deals with this type of crime by applying traditional laws relating to ordinary crimes, not cyber-related laws. Spring Land does not have any regulations or laws specifically relating to privacy, data protection or human rights.

*“There are no laws related to protect systems and data” NCSA team.*



**Figure 2.** Results of all CCMM dimensions.

In addition, law enforcement along with the investigation and prosecution of cybercrime services in Spring Land face a shortage of skills to handle cybercrime cases. Moreover, there is no national mechanism to report or disclose cyber related crime or vulnerabilities. Also, there are no specific courts dealing with digital crime and no training is provided to build capacity in this particular dimension.

### 5.2.5. Standards, organisations and technology indicators (D5)

Throughout the discussion, all of the participants agreed that Spring Land is at the start-up stage in terms of this dimension. There are no cybersecurity standards that have been adapted to procurement and software development in the government sector. As explained by the NCSA team, there has been an attempt to start the process of implementing international standards but there is a shortage of skilled people and financial resources. There is no national agency or framework to monitor the implementation of standards and minimal acceptable practices in the government sector.

In addition, there is a lack of research centres in this field and poor co-operation between the public and private sectors in terms of training and the development of skills. As mentioned by the participants in the discussion of Dimension 1, not all sectors have a disaster recovery plan or a business continuity plan. All of the participants pointed out that the government does not have a plan to manage, monitor or evaluate national infrastructure resilience.

## 6. Conclusion and future work

The current paper has examined the core features of Spring Land's cybersecurity to demonstrate the typical situation faced by countries in transition phase and how best to address such matters. The observations help to improve our grasp of the capacity of cybersecurity in Spring Land and provide a foundation for a National Cybersecurity Capacity Framework (NCCBF) in countries that are transitioning stages. The IM approach yielded a series of problem statements and objectives that can be applied to

enhance management processes in similar cases. Be that as it may, it is apparent that additional validation is needed for the results obtained and future research should select data that will enable the results to be generalised. In addition, future research should apply the UML and IDEF0 modelling methods so that the ISM can be decomposed into functional models to develop the NCCBF. Using data from advanced countries in which the CCMM is already operational in addition to a range of practices and standards, the framework will make it possible for transitioning countries to overhaul their existing cybersecurity arrangements by initiating strategies capable of realising a desirable outcome.

## References

- [1] M. McCrabb, "Rough Waters," *Naval War College Review*, vol. 70, pp. 141-145, 2017.
- [2] M. L. Cook, R. H. Dorff, D. Jablonsky, M. G. Roskin, R. C. Nation, G. Marcella, *et al.*, *US Army War College Guide to Strategy*: Strategic Studies Institute, 2001.
- [3] DeRouen Jr, Karl Goldfinch, and Shaun, "What makes a state stable and peaceful? good governance, legitimacy and legal-rationality matter even more for low-income countries," *Civil Wars*, vol. 14, pp. 499-520, 2012.
- [4] D. W. Brinkerhoff, "Rebuilding governance in failed states and post-conflict societies: core concepts and cross-cutting themes," *Public Administration and Development: The International Journal of Management Research and Practice*, vol. 25, pp. 3-14, 2005.
- [5] S. Guo and G. A. Stradiotto, *Democratic transitions: Modes and outcomes*: Routledge, 2014.
- [6] Strachan and Anna, "Factors affecting success or failure of political transitions," Institute of Development Studies, K4D Helpdesk Report. Brighton, UK2017.
- [7] M. I. Wilson and K. E. Corey, "The role of ICT in Arab spring movements," *Netcom. Réseaux, communication et territoires*, pp. 343-356, 2012.
- [8] GCSCC. (2017, 25/10/2016). *Cybersecurity Capacity Maturity Model for Nations (CMM)* Available: [https://www.sbs.ox.ac.uk/cybersecurity-capacity/system/files/CMM%20Version%201\\_2\\_0.pdf](https://www.sbs.ox.ac.uk/cybersecurity-capacity/system/files/CMM%20Version%201_2_0.pdf)
- [9] L. P. Muller, "Cyber security capacity building in developing countries: challenges and opportunities," 2015.
- [10] M. A. B. Naseir, H. Dogan, E. Apeh, C. Richardson, and R. Ali, "Contextualising the National Cyber Security Capacity in an Unstable Environment: A Spring Land Case Study," in *World Conference on Information Systems and Technologies*, 2019, pp. 373-382.
- [11] P. Pawlak, "Capacity building in cyberspace as an instrument of foreign policy," *Global Policy*, vol. 7, pp. 83-92, 2016.
- [12] E. Tamarkin, "The AU's cybercrime response: A positive start, but substantial challenges ahead," 2015.
- [13] R. Azmi, W. Tibben, and K. T. Win, "Review of cybersecurity frameworks: context and shared concepts," *Journal of Cyber Policy*, vol. 3, pp. 258-283, 2018.
- [14] A. C. Tagert, "Cybersecurity challenges in developing nations," 2010.
- [15] I. Atoum, A. Otoom, and A. A. Ali, "A holistic cyber security implementation framework," *Information Management & Computer Security*, 2014.
- [16] NCSA. (2013). *The National Cyber Security Authority (NCSA)*
- [17] Warfield, John N. and A. R. Cárdenas, *A handbook of interactive management*: Iowa State Press, 2002.
- [18] F. R. Janes, "Interactive Management: Framework, Practice, and Complexity," 1995, pp. 51-60.
- [19] A. Tong, P. Sainsbury, and J. Craig, "Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups," *International journal for quality in health care*, vol. 19, pp. 349-357, 2007.
- [20] O. Doody, E. Slevin, and L. Taggart, "Focus group interviews. Part 3: analysis," *British Journal of Nursing*, vol. 22, pp. 266-269 4p, 2013.
- [21] J. Kitzinger, "Qualitative research: introducing focus groups," *Bmj*, vol. 311, pp. 299-302, 1995.
- [22] P. Checkland, *Soft systems methodology : a 30-year retrospective ; Systems thinking, systems practice*: Chichester : John Wiley, c1999., 1999.
- [23] E. Trist, "The socio-technical perspective: The evolution of socio-technical systems as a conceptual framework and as an action research paradigm," ed: New York: Wiley & Sons, 1981.

# Algorithm Selection for Machine Learning Classification: An Application of the MELCHIOR Multicriteria Method

Igor Pinheiro de Araújo COSTA<sup>a,d,1</sup>, Marcio Pereira BASÍLIO<sup>b</sup>, Sérgio Mitihiro do Nascimento MAÊDA<sup>a,d</sup>, Marcus Vinícius Gonçalves RODRIGUES<sup>a</sup>, Miguel Ângelo Lellis MOREIRA<sup>a,d</sup>, Carlos Francisco Simões GOMES<sup>a</sup> and Marcos dos SANTOS<sup>c,d</sup>

<sup>a</sup>Federal Fluminense University, Brazil

<sup>b</sup>Military Police of the State of Rio de Janeiro, Brazil

<sup>c</sup>Military Institute of Engineering, Brazil

<sup>d</sup>Naval Systems Analysis Center, Brazil

**Abstract.** This paper aims to select an algorithm for the Machine Learning (ML) classification task. For the proposed analysis, the Multi-criteria Decision Aid (MCDA) *Méthode d'ELimination et de CHOix Incluant les relations d'ORDre* (MELCHIOR) method was applied. The experiment considered the following criteria as relevant: Accuracy, sensitivity, and processing time of the algorithms. The data used refers to the intention of buying on the Internet and the purpose is to predict whether the customer will finalize a particular purchase. Among various MCDA techniques available, MELCHIOR was chosen to support the decision-making process because this method provides the evaluation of alternatives without the need to elicit the weights of the criteria. As a result, the Gradient Boosting Decision Tree algorithm has been selected as the most suitable for the ML classification task.

**Keywords.** Multi-criteria Decision Analysis (MCDA), Machine Learning, Outranking, MELCHIOR.

## 1. Introduction

The growth of the "data-driven" culture opens space for decision-making and Machine Learning (ML) techniques. The demand for methods and models that generate quality information for academic and professional purposes is growing. The ML grew as a subfield of Artificial Intelligence (AI), developing an important role in research and day-to-day [1].

The classification, a common task of ML, aims to predict binomial or multinomial categorical values. Some examples are product purchase and service cancellation predictions, as well as fraud detection and default risk.

The selection of algorithms in ML can be understood as a problem of multiple alternatives and criteria. Therefore, the purpose in this paper is to explore this possible

---

<sup>1</sup> Corresponding Author, Igor Pinheiro de Araújo COSTA, Department of Production Engineering, Federal Fluminense University, Niterói, Brazil; E-mail: [costa\\_igor@id.uff.br](mailto:costa_igor@id.uff.br).

interaction between multicriteria analysis and AI. In this context, the Multicriteria Decision Analysis or Aiding (MCDA) methods aim to help in understanding the decision-making process and choosing alternatives in front of multiple criteria [2,3].

Multicriteria methods consider value judgments and not only technical issues [4], and tend to be increasingly adopted to address the real-world construction problems [5]. These methods have been used to support the decision-making process in several recent complex problems, as presented in [6–12].

Regarding the application of MCDA methods in ML-related problems, the literature presents several cases, such as in the analysis of human decision-making through learning preferences [13]; in a case-clearance procedure for COVID-19 [14]; evaluation of emergency prediction models [15]; supplier performance classification using the Random Forest ML algorithm [16]; selection of classification algorithms for financial risk forecasts [17]; and for propose a new Support Vector Machine (SVM) model based on density weight for binary Class Imbalance Learning CIL problem [18]. Additionally, an improved 2-norm-based density-weighted least squares SVM for binary CIL (IDLSSVM-CIL) is also proposed to increase the training speed of DSVM-CIL.

Given the importance of classification for the success of organizations, the goal of this paper is to select an algorithm for the ML classification task, by applying the *Méthode d'ELimination et de Choix Incluant les relations d'ORDre* (MELCHIOR) MCDA model.

In this research, the decision-makers (DM) reported a difficulty in evaluating and establishing the weights of the criteria. In these cases, the MELCHIOR method has good adherence, since it provides the evaluation of alternatives without the need to elicit the weights of the criteria, besides not considering interaction between them. Therefore, in this paper we chose to apply this outranking method as a tool to support decision making.

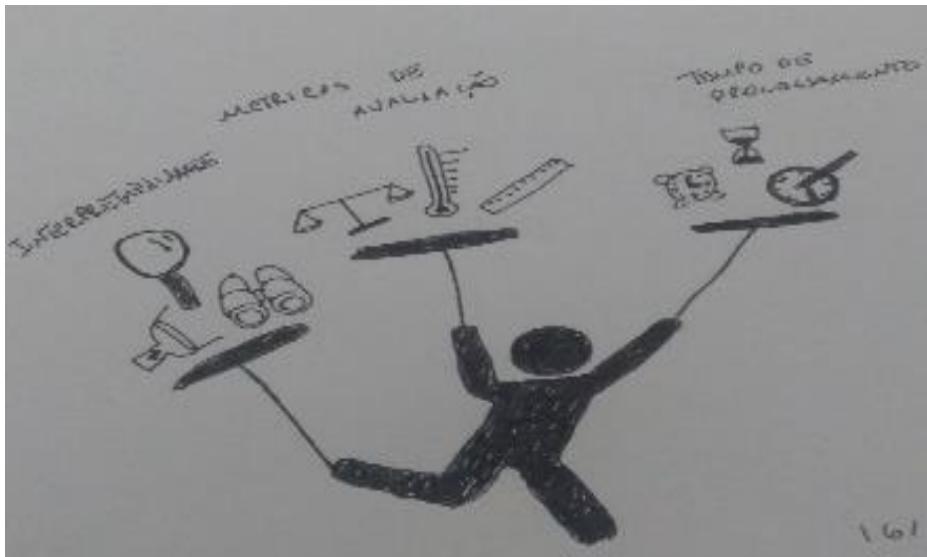
This work is divided into 4 sections besides this introduction. Section 2 discusses the understanding of the problematic situation, with the definition of the criteria and alternatives that make up the proposed case study. Section 3 presents the background of the MELCHIOR method, while section 4 addresses the methodology and the application of the MELCHIOR method to support the decision-making process in the proposed case study. Finally, section 5 concludes this study.

## 2. Problem Structuring

To help understand the problem, in this article we applied a Problem Structuring Method (PSM) established in the literature – the Soft Systems Methodology (SSM) [19]. Among the most commonly used and consolidated methods in the literature, SSM has been explored in a variety of research fields, as well as serves equally diverse practical interests [20]. According to [19], SSM presents seven stages of application, two of which were addressed in this article for structuring the problem: 1: exploring an unstructured problematic situation; and 2: express it.

In the first stage, the brainstorming technique was used by the authors to demonstrate the group's perceptions about all possible information, without interference or judgment to define the problem. In the second stage, a rich picture was constructed (Figure 1), which has great value as a starting point in the exploratory analysis of the problem [21]. The rich picture is a simple SSM tool, extremely useful for opening the discussion around individual perceptions toward a broad view of the different issues affecting the situation.

They are created freely and unstructured to capture the participants' interpretation of a real situation [19,20].



**Figure 1.** Rich Picture.

The purpose of the case study is to choose a classifier algorithm to predict whether the individual on the internet will finalize the purchase (positive class) or not (negative class). The rich picture portrays the attempt to balance the criteria in choosing the algorithm. Regarding the quantitative criteria, while evaluation metrics reflect the algorithm's effectiveness, processing time is a measure of performance. The metrics of evaluation accuracy and sensitivity are based on the confusion matrix (Table 1).

**Table 1.** Confusion matrix.

	Positive	Negative
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Positive (TP)

Accuracy measure assertiveness when predicting positive class and its formula is  $TP/(TP+FP)$ . Sensitivity, on the other hand, can achieve and classify positive scans, its formula is  $TP/(TP+FN)$ . For the processing time, the average time of 3 runs was considered. In addition, the following algorithms were considered as alternatives:

- K-nearest neighbor (KNN), which aims to find the nearest Neighbor K of a new sample and perform the prediction based on them;
- Support Vector Machine (SVM), which works to find the hyperplane that best separates the training base, maximizing the distance between the hyperplane and the points closest to it, in order to avoid wrong classifications of new samples;
- Random Forest (RF): a set of decision trees built into data samples; and
- Gradient Boosting Decision Tree (GBM): which performs sequential training, developing new models from previous model errors.

### 3. The MELCHIOR method

For the establishment of preference relationships, the MELCHIOR method, proposed by [22], establishes three fundamental situations of comparison between alternatives:

I – Weak Preference (q): There are clear and positive reasons that do not imply a strict preference in favor of one (well defined) of the two actions, but these reasons are insufficient to assume a strict preference in favor of another, or the indifference between them [23];

II - Strict preference (p): There are clear and positive reasons that justify a significant preference in favor of one (well defined) of the two actions;

III – Veto (v): Limit defined for each criterion that sets a value for the difference  $g_j(b) - g_j(a)$  (difference in relation to criterion  $j$  and discordant of the aSb statement), from which the proposition aSb will not be accepted [24].

In the MELCHIOR method, the basic information is a family F of pseudocriteria, that is, criteria  $g_j$  with indifference threshold  $q_j$  and a preference threshold  $p_j$  ( $p_j > q_j \geq 0$ ) in such a way that,  $\forall j \in J$  and  $\forall a, b \in A$  [25]:

- $a$  is strictly preferable to  $b$  ( $aP_jb$ ) in relation to  $g_j$  if  $g_j(a) > g_j(b + p_j[g_j(b)])$ ;
- $a$  is weakly preferable to  $b$  ( $aQ_jb$ ) in relation to  $g_j$  if  $g_j(b) + p_j[g_j(b)] \geq g_j(a) > g_j(b) + q_j[g_j(b)]$ ;
- $a$  and  $b$  are indifferent ( $aI_jb$ ) if there is no strict or weak preference between them.

In the MELCHIOR method no weight is assigned to the criteria. A binary relationship M in F is defined in such a way that  $g_iMg_j$  means that "criterion  $g_j$  is as important as criterion  $g_i$ " [25].

In order to obtain the comprehensive outranking relationship aSb, Leclercq [22] proposed a particular form of analysis, in which the criteria for and against the outranking relationship are evaluated to verify agreement if there is no situation of disagreement. That is, no criterion  $g_j$  of F exists such that  $g_j(b) > g_j(a) + v_j$ , where  $v_j$  is a veto threshold for criterion  $g_j$  (absence of disagreement).

In this method, a criterion  $g_j \in F$  is said to be in favor of the aSb outranking if one of the following situations are verified:

- $aP_jb$  (strict marginal preference of  $a$  in relation to  $b$ ) (1st condition);
- $aP_jb$  or  $aQ_jb$  (strict or weak marginal preference of  $a$  in relation to  $b$ ) (2nd condition);
- $g_j(a) > g_j(b)$  (3rd condition).

A criterion  $g_j \in F$  is said to be against the aSb outranking relationship if one of the following situations are verified:

- $bP_ja$  (strict marginal preference of  $b$  over  $a$ ) (1st condition);
- $bP_ja$  or  $bQ_ja$  (strict or weak marginal preference of  $b$  over  $a$ ) (2nd condition);
- $g_j(b) > g_j(a)$  (3rd condition).

The analysis of agreement of the outranking relationship aSb, for  $a, b \in A$ , is made verifying whether the family of G criteria in favor of this relationship "masks" the family of h criteria that are against the relationship aSb [25]. These subsets of criteria are compared only using the binary relationship M in F. It is said that a subset G of criteria "masks" a subset H of criteria ( $G, H \subset F, F \cap G = \emptyset$ ) if, for each criterion  $g_i$  of H, there is a criterion  $g_j$  of G such that:

- $g_jMg_i$  (1st condition); or
- $g_jMg_i$  or not ( $g_jMg_i$ ) (2nd condition).

Where the same criterion  $g_j$  of G is allowed to mask at most one criterion of H. Leclercq [22] explains that by choosing two appropriate combinations of the above conditions, the first being more rigorous than the second, and verifying the agreement and absence of disagreement, a strong or weak outranking relationship can be constructed respectively (Table 2).

**Table 2.** Establishment of outranking relationships.

Relationship	Conditions
$aP_j^+b$	If $g_j(a) > g_j(b) + p_i$
$aQ_j^+b$	If $g_j(a) > g_j(b) + p_i$ and $g_j(a) > g_j(b) + q_i$
$aI_j^+b$	If $g_j(a) > g_j(b) + q_i$ and $g_j(a) > g_j(b)$
$aE_jb$	If $g_j(a) = g_j(b)$
$aP_j^-b$	If $bP_j^+a$
$aQ_j^-b$	If $bQ_j^+a$
$aI_j^-b$	If $bI_j^+a$

For the establishment of strong and weak outranking relationships between alternatives, Leclercq [22] defines:

I) Strong outranking ( $S_F$ ): For an alternative  $a$  to present a strong overcoming relationship over  $b$ , necessarily:

- There are no criteria for which  $b$  is strictly preferable to  $a$ ;
- Criterion  $i$  for which  $b$  is weakly preferable to  $a$  must be masked by more important criteria for which  $A$  enjoys strict preference.

II) Weak outranking ( $S_f$ ): For an alternative  $a$  to present a weak outranking over  $b$ , it is necessary that the criteria  $i$  for which  $b$  has the advantage must be masked by criteria  $j$  at least as important in favor of  $a$ .

Finally, it is emphasized that, when applying the MELCHIOR method, no possibility of interaction between criteria is considered, since the outranking relationships are constructed by analyzing, one by one, the criteria for and against the relation  $aSb$  [25].

#### 4. Case Study

The proposal is to use the MELCHIOR method to select a classifier to predict whether or not the individual will make an online purchase. The data contains information about the date, access behavior, and characteristics of the individual. The database used is part of the UCI Machine Learning repository and has 12,330 observations, 18 attributes and approximately 16% of the data are related to consumers who have completed the purchase. So, it is an unbalanced base.

At first, the sets of alternatives and criteria for structuring the problem were inserted: K-nearest neighbor (KNN), Support Vector Machine (SVM), Random Forest (RF) and Gradient Boosting Decision Tree (GBM) as alternatives, as well as accuracy, sensitivity, and processing time as criteria. Table 3 shows the performances of alternatives in the light of the established criteria, as well as the strict (p), weak (q) and veto (v) preference thresholds, established together with specialists in ML.

**Table 3.** Performance Matrix.

Alternatives	Accuracy	Sensitivity	Processing time (s)
KNN	39.6	19.8	4.5
SVM	71.5	34.6	7.5
RF	67.2	62.6	5.2
GBM	66.3	61	4.7

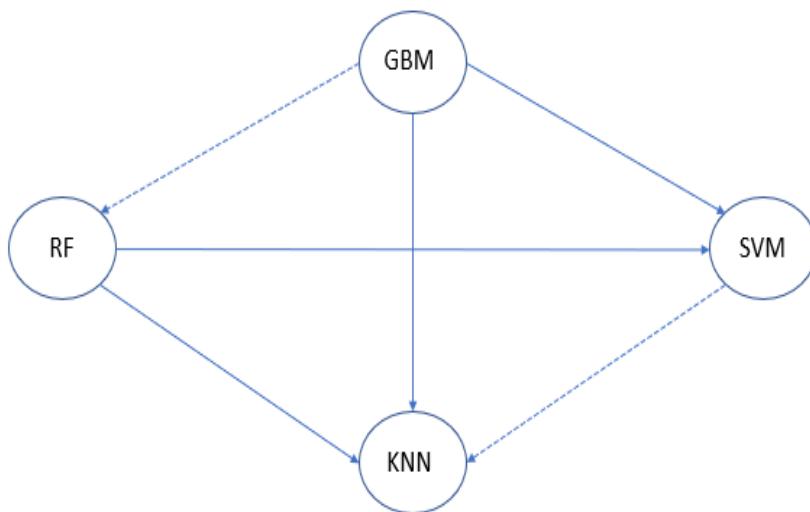
q	10	13	0.5
p	20	20	1
v	40	50	4

We emphasize that Accuracy and Sensitivity are maximizing criteria, while processing time, minimization. After defining the alternatives, criteria, preference thresholds and veto, the MELCHIOR method can be applied. Table 4 illustrates the outranking relationships, based on [22]:

**Table 4.** Establishment of outranking relationships.

Pairwise evaluation	Accuracy	Sensitivity	Processing time (s)	Relationship
KNN/SVM	P-	Q-	P <sup>+</sup>	SVM S <sub>F</sub> KNN
KNN/RF	P-	P-	Q <sup>+</sup>	RF S <sub>F</sub> KNN
KNN/GBM	P-	P-	I	GBM S <sub>F</sub> KNN
SVM/RF	I	P-	P-	RF S <sub>F</sub> SVM
SVM/GBM	I	P-	P-	GBM S <sub>F</sub> SVM
RF/GBM	I	I	Q-	GBM S <sub>f</sub> RF

Where S<sub>F</sub> represents strong outranking and S<sub>f</sub> illustrates a weak outranking between alternatives. Thus, with the relationships established between the alternatives, it is possible to generate an outranking graph (Figure 2).



**Figure 2.** Graph representing the outranking relationships between the alternatives.

The origin of the edges represents the alternative that outranks, while the target, the overqualified. Dotted arrows depict weak outranking relationships (S<sub>f</sub>), while continuous ones symbolize strong relationships (S<sub>F</sub>). Analyzing the graph, we obtain the final outranking ratio, according to [22]:

- GBM } RF } SVM } KNN.

We emphasize that, in this research, the same symbols (}) used by [22] were applied to represent the outranking relationships, that is, the relation GBM } RF means that the first alternative outranks the second one.

In view of the above, the GBM can be considered as the most indicated alternative to be selected as an algorithm for the ML classification task. Analyzing the reasons that justify this choice, it is observed that the GBM and RF alternatives present good

performances in all the analyzed criteria, with relationships of indifference in the criteria accuracy and sensitivity.

We observe that the criterion that defined the choice of GBM was the processing time, defining the weak outranking relationship between the two best alternatives in favor of the GBM algorithm.

## 5. Conclusions

The ML classification task contributes to the prediction and understanding of results in various sectors. The MELCHIOR method effectively supported the decision to choose a classifier algorithm. The chosen alternative, Gradient Boosting Decision Tree, was selected as the most indicated algorithm, and its result was justified by a good performance in all evaluated criteria, which provides credibility to the result achieved.

The application of the MELCHIOR method occurred in a context in which the decision-makers claimed great difficulty in eliciting the weights of the criteria, which justifies the option for a multicriteria method that does not present interactions and that allows the analysis of alternatives without assigning weights to the criteria.

In view of the above, it was clear that the methodology presented in this paper can be used to solve problems of various types, considering that it presents a simple, flexible, reliable and fast methodology. Future work could address comparative analyses or hybrid modeling of the MELCHIOR method with other MCDA tools to support high-level decision-making on tactical, operational, strategic, and political level issues.

As a limitation of this study, we highlight that, among different types of LM algorithms, only four were evaluated. Future research could address more ranking models, as well as a greater number of criteria for analyzing systems..

## References

- [1]. Raschka S, Mirjalili V. Python Machine Learning: Machine Learning and Deep Learning with Python. Scikit-Learn, TensorFlow Second Ed ed. 2017;
- [2]. Oliveira AS, Gomes CFS, Clarkson CT, Sanseverino AM, Barcelos MRS, Costa IPA, et al. Multiple Criteria Decision Making and Prospective Scenarios Model for Selection of Companies to Be Incubated. Algorithms [Internet]. 2021 Mar 30;14(111). Available from: <https://www.mdpi.com/1999-4893/14/4/111>
- [3]. Filip FG, Zamfirescu C-B, Ciurea C. Computer-supported collaborative decision-making. Springer; 2017.
- [4]. Santos M dos, Quintal RS, Paixão AC da, Gomes CFS. Simulation of Operation of an Integrated Information for Emergency Pre-Hospital Care in Rio de Janeiro Municipality. Procedia Comput Sci [Internet]. 2015;55:931–8. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1877050915015860>
- [5]. Zhu X, Meng X, Zhang M. Application of multiple criteria decision making methods in construction: a systematic literature review. J Civ Eng Manag. 2021;27(6):372–403.
- [6]. Costa IP de A, Maêda SM do N, Teixeira LFH de S do B, Gomes CFS, Santos M dos. Choosing a hospital assistance ship to fight the Covid-19 pandemic. Rev Saude Publica [Internet]. 2020;54. Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85090141296&doi=10.11606%2FS1518-8787.2020054002792&partnerID=40&md5=90355a4a86a1b09b1add8956ace15019>
- [7]. Costa IP de A, Sanseverino AM, Barcelos MR dos S, Belderrain MCN, Gomes CFS, Santos M dos. Choosing flying hospitals in the fight against the COVID-19 pandemic: structuring and modeling a complex problem using the VFT and ELECTRE-MOr methods. IEEE Lat Am Trans [Internet]. 2021 Jun;19(6):1099–106. Available from: <https://ieeexplore.ieee.org/document/9451257/>
- [8]. Santos M dos, Costa IP de A, Gomes CFS. Multicriteria decision-making in the selection of warships: a new approach to the AHP method. Int J Anal Hierarchy Process [Internet]. 2021 May 19;13(1). Available from: <https://ijahp.org/index.php/IJAHP/article/view/833>

- [9]. Tenório FM, dos Santos M, Gomes CFS, Araujo J de C. Navy Warship Selection and Multicriteria Analysis: The THOR Method Supporting Decision Making. In: International Joint conference on Industrial Engineering and Operations Management [Internet]. Springer; 2020. p. 27–39. Available from: [http://link.springer.com/10.1007/978-3-030-56920-4\\_3](http://link.springer.com/10.1007/978-3-030-56920-4_3)
- [10]. Moreira MÁL, Costa IP de A, Pereira MT, dos Santos M, Gomes CFS, Muradas FM. PROMETHEE-SAPEVO-M1 a Hybrid Approach Based on Ordinal and Cardinal Inputs: Multi-Criteria Evaluation of Helicopters to Support Brazilian Navy Operations. Algorithms [Internet]. 2021 Apr 27;14(5):140. Available from: <https://www.mdpi.com/1999-4893/14/5/140>
- [11]. Maêda SM do N, Costa IP de A, Castro Junior MAP, Fávero LP, Costa AP de A, Corriça JV de P, et al. Multi-criteria analysis applied to aircraft selection by Brazilian Navy. Production. 2021;31:1–13.
- [12]. de Almeida IDP, Corriça JV de P, Costa AP de A, Costa IP de A, Maêda SM do N, Gomes CFS, et al. Study of the Location of a Second Fleet for the Brazilian Navy: Structuring and Mathematical Modeling Using SAPEVO-M and VIKOR Methods. ICPR-Americas 2020 Commun Comput Inf Sci [Internet]. 2021;1408:113–24. Available from: [https://link.springer.com/10.1007/978-3-030-76310-7\\_9](https://link.springer.com/10.1007/978-3-030-76310-7_9)
- [13]. Guo M, Zhang Q, Liao X, Chen FY, Zeng DD. A hybrid machine learning framework for analyzing human decision-making through learning preferences. Omega (United Kingdom) [Internet]. 2021;101. Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85083426067&doi=10.1016%2Fj.omega.2020.102263&partnerID=40&md5=00f6a289b864d90c453e7f59a9607023>
- [14]. Guhathakurata S, Saha S, Kundu S, Chakraborty A, Banerjee JS. South Asian Countries are Less Fatal Concerning COVID-19: A Fact-finding Procedure Integrating Machine Learning & Multiple Criteria Decision-Making (MCDM) Technique. J Inst Eng Ser B [Internet]. 2021; Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85102293277&doi=10.1007%2Fs40031-021-00547-z&partnerID=40&md5=6a4c42319e62df3fba2beb4133db9beb>
- [15]. Mousavi MM, Lin J. The application of PROMETHEE multi-criteria decision aid in financial decision making: Case of distress prediction models evaluation. Expert Syst Appl [Internet]. 2020;159. Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85086503565&doi=10.1016%2Fj.eswa.2020.113438&partnerID=40&md5=18af8984d83302f6c387c04c45f83d0a>
- [16]. Wilson VH, Prasad ANS, Shankharan A, Kapoor S, Rajan JA. Ranking of supplier performance using machine learning algorithm of random forest. Int J Adv Res Eng Technol [Internet]. 2020;11(5):298–308. Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85086229513&doi=10.34218%2FIJARET.11.5.2020.031&partnerID=40&md5=1ae77633ce02dc30a09a23add55c047d>
- [17]. Song Y, Peng Y. A MCDM-based evaluation approach for imbalanced classification methods in financial risk prediction. IEEE Access. 2019;7:84897–906.
- [18]. Hazarika BB, Gupta D. Density-weighted support vector machines for binary class imbalance learning. Neural Comput Appl [Internet]. 2021;33(9):4243–61. Available from: <https://doi.org/10.1007/s00521-020-05240-8>
- [19]. Checkland PB. Systems Theory. Syst Pract. 1981;
- [20]. Rose J. Soft systems methodology as a social science research tool. Syst Res Behav Sci Off J Int Fed Syst Res. 1997;14(4):249–58.
- [21]. Checkland P. Systems thinking, systems practice: includes a 30-year retrospective. Journal-Operational Res Soc. 2000;51(5):647.
- [22]. Leclercq J. Propositions d’extension de la notion de dominance en présence de relations d’ordre sur les pseudo-critères: la méthode MELCHIOR. JORBEL-Belgian J Oper Res Stat Comput Sci. 1984;24(1):32–46.
- [23]. Gomes CFS, Gomes LFAM, Maranhão FJC. Decision analysis for the exploration of gas reserves: merging TODIM and THOR. Pesqui Operacional. 2010;30(3):601–17.
- [24]. Gomes L, Gomes CFS. Princípios e métodos para a tomada de decisão: Enfoque multicritério. São Paulo: Atlas; 2019.
- [25]. Martel J-M, Matarazzo B. Other outranking approaches. In: Multiple Criteria Decision Analysis. Springer; 2016. p. 221–82.

# Investments in Times of Pandemics: An Approach by the SAPEVO-M-NC Method

Sérgio Mitihiro do Nascimento MAÊDA<sup>a,d,1</sup>, Marcio Pereira BASÍLIO<sup>b</sup>, Igor Pinheiro de Araújo COSTA<sup>a,d</sup>, Miguel Ângelo Lellis MOREIRA<sup>a,d</sup>, Marcos dos SANTOS<sup>c,d</sup>, Carlos Francisco Simões GOMES<sup>d</sup>, Isaque David Pereira de Almeida<sup>a</sup> and Arthur Pinheiro de Araújo Costa<sup>d</sup>

<sup>a</sup>*Federal Fluminense University, Brazil*

<sup>b</sup>*Military Police of the State of Rio de Janeiro, Brazil*

<sup>c</sup>*Military Institute of Engineering, Brazil*

<sup>d</sup>*Naval Systems Analysis Center, Brazil*

**Abstract.** The consequences of the pandemic caused by the new coronavirus in the most diverse sectors of the Brazilian economy, are overwhelming, and its effects are still difficult to measure completely. There are several possible scenarios being considered, such as prolonged depression, “U” or “V” recovery. Due to such volatility, risks and uncertainties, the investor, before investing, must carefully analyze the alternatives available in the market. Given the above, this article aims to propose different ways of distributing a financial portfolio, considering five investment funds, which were evaluated in the light of five criteria, by two investors who work in the financial market. Therefore, the SAPEVO-M-NC multicriteria decision aid method was used to evaluate the alternatives, as well as their composition in the investment portfolios. The adoption of the methodology made it possible to carry out the distribution of the portfolio in a clear and consistent way, showing itself as an efficient practical tool for the proposed approach.

**Keywords.** Pandemic, COVID-19, Investments, Multicriteria decision aid

## 1. Introduction

The consequences of the new coronavirus pandemic, caused by the new coronavirus in the most diverse sectors of the Brazilian economy, are overwhelming and its effects in the medium and long term are still difficult to measure [1][2]. Uncertainties result in the flight of foreign capital and influence investments by national organizations in companies in more developed countries and with varied and consolidated corporate governance structures, thus seeking to mitigate investment risks [3]. During the decision-making process, the evaluator will come across different information, often conflicting with each other[4][5][6]. To provide greater transparency to the process, as well as to mitigate possible errors involved, the research sources must be exhaustively consulted,

---

<sup>1</sup> Corresponding Author, Sérgio Mitihiro do Nascimento Maêda, Federal Fluminense University - UFF, 24210-240, Niterói, Brazil; E-mail: [sergiomnmaeda@gmail.com](mailto:sergiomnmaeda@gmail.com).

This research was financially supported by Military Police of the State of Rio de Janeiro.

using reliable publications, which must be refined according to the research objective, providing the evaluator with the tools needed for a more accurate decision-making [7][8][9].

This article aims to propose the distribution of a financial portfolio, considering five investment funds, which were evaluated considering five criteria, by two investors who operate in the Brazilian financial market. Therefore, the SAPEVO-M-NC (Simple Aggregation of Preferences Expressed by Ordinal Vectors –Multi Decision Makers Non-Compensatory) multicriteria decision aid (MCDA) method was used to evaluate the alternatives, as well as their distribution in the investment portfolios. The adoption of the methodology made it possible to carry out the distribution of portfolios in a clear, robust way and with relatively low cognitive effort on the part of decision makers (DM).

## 2. Materials and methods

In this section, the alternatives, criteria, and the SAPEVO-M-NC method will be presented. To assess the importance of the criteria and the performance of the alternatives, interviews were conducted with two professionals ( $DM_1$  and  $DM_2$ ) who work in the financial market. It is noteworthy that these are not specialists in investments, having in this activity a means of increasing their earnings in the medium and long term. Asked about their investment profiles,  $DM_1$  declared itself as a “moderate” profile investor, while  $DM_2$  declared itself as an investor between “moderate and bold” profiles.

### 2.1. Alternatives

The alternatives indicated to compose the investment options in this article were selected from a Brazilian investment broker with great influence on the financial market.

Investment A<sub>1</sub> (FJPF): Post-fixed fixed income investment fund that invests at least 80% of the portfolio in federal government bonds and/or assets with low market credit risk. In this fund, when interest rates are high, yields tend to rise and when interest rates fall, profitability tends to show the same direction.

Investment A<sub>2</sub> (FJPR): Fixed-income investment fund, with exposure to real and nominal interest in Brazil and the possibility of operating in the international market, seeking medium and long-term yields above the CDI in different scenarios.

Investment A<sub>3</sub> (FM): Multimarket investment fund exposed to various types of assets, such as interest, government bonds, foreign exchange, domestic and foreign shares, and variable income, offering more freedom for the manager to assemble the investment strategy, exploring opportunities in different scenarios.

Investment A<sub>4</sub> (FA): Stock fund that has exposure to a portfolio composed of stocks that seek to monitor the performance of the Ibovespa (Brazil stock exchange), being diversified and representing the behavior of the stock market.

Investment A<sub>5</sub> (FC): Fund with exposure to financial assets related to the US dollar, which aims to reduce the risks of exchange variation, or for investors seeking an investment linked to the variation of the US dollar.

### 2.2. Criteria

Criterion C<sub>1</sub> – Profitability: indicates the percentage of return on invested capital, being a monotonic profit criterion, that is, the higher the better.

Criterion C<sub>2</sub> - Risk: represents the investment security levels, which will be considered market and credit. In this article, this criterion was classified using a 4-point scale, ranging from grade 0, which represents zero risk, to grade 5, high risk.

Criterion C<sub>3</sub> - Liquidity: ease of investment being “transformed into cash” without significant losses in its value. It should be noted that some of the investments have a previously established withdrawal date.

Criterion C<sub>4</sub> - Volatility: Defined as the variation, over time, of the conditional variance of a time series. The perception of damage generated by the loss is 2.25 times greater than the sense of benefit produced by the gain, thus, this criterion will be considered of a monotonic cost nature, being the worse, the more expressive.

Criterion C<sub>5</sub> - Initial application: Minimum amount required by the financial institution so that the investor can join the investment alternative. A high initial contribution acts preventing the access of less capitalized investors to the considered asset; thus, this criterion will also be considered of a monotonic cost nature.

The table 1 presents the data for the selected investments.

**Table 1.** Data on evaluated investments

Investment	Profitability in the last 12 months	Average profitability	Risk	Liquidity	Volatility in the last 12 months	Annualized volatility	initial application
A <sub>1</sub> (FJPF)	2,40%	5,45%	1	D	0,31 %	0,24 %	R\$ 1,00
A <sub>2</sub> (FJPR)	4,35%	3,53%	2	1+D	0,93 %	0,97 %	R\$ 500,00
A <sub>3</sub> (FM)	7,39%	21,11%	3	17+D	1,21 %	1,45 %	R\$ 50.000,00
A <sub>4</sub> (FA)	18,38%	18,93%	4	30+D	40,40 %	25,2 %	R\$ 15.000,00
A <sub>5</sub> (FC)	25,30%	12,03%	4	1+D	19,7 %	15,9 %	R\$ 1,00

### 2.3. The SAPEVO-M-NC Method

Derived from the SAPEVO-M ordinal method [10], the SAPEVO-M-NC method [11], consists of an ordinal method, non-compensatory nature, which aims at the problem of ordering ( $P\gamma$ ), and with the possibility of acting by multiple decision makers. In the method, the evaluation of the performance of the alternatives is carried out directly, with no need to carry out parity comparisons between the alternatives to obtain the modeling of preferences between them, resulting in a substantial reduction in the cognitive effort on the part of the DMs. The method also allows evaluating the rate of over classification between the alternatives. This information enables not only the DM to verify the dominance relationship, but also to identify, in percentage terms, how much one alternative outperforms the other [11].

Unlike Classic Decision Theory, which basically considers two supposedly transitive preference relations, designated by Indifference (I) and by Strict Preference (P), this method is based on the Fundamental System of Preference Relations (SFRP), also incorporating the weak preference relationship (Q). According to [12], there are several reasons why researchers seek to avoid the type of modeling that is based on the axiom of complete comparability and transitivity between alternatives, among which, can be mentioned the fact that the decision maker does not have all the information that allows him to choose one of the alternatives and by forcing a relationship of strict preference, or indifference, could lead to arbitrary and inconsistent errors. After the DMs establish the criteria and alternatives, the method can be divided into six steps:

Step 1 - Ordinal transformation of the preferences of each DM, in each criterion, which are added at the end of this step, giving rise to a vector ( $V_i$ ), representing the weights of the criteria [10].

Let  $c_i$  and  $c_j$  be two criteria within a set of criteria  $C = \{c_1, c_2, c_3, \dots, c_i, c_j\}$ . The degree of preference between them is given by:  $\delta(c_i, c_j)$ . Table 2 shows the relationship of relative importance between the criteria:

**Table 2.** Relative importance between the criteria

Scale 1	Verbal representation	Scale 2
$\delta(c_i, c_j) \ll 1 \leftrightarrow c_i << c_j$	Absolutely less important	-3
$\delta(c_i, c_j) \prec 1 \leftrightarrow c_i < c_j$	Much less important	-2
$\delta(c_i, c_j) \prec 1 \leftrightarrow c_i < c_j$	Less important	-1
$\delta(c_i, c_j) = 1 \leftrightarrow c_i \approx c_j$	Equally important	0
$\delta(c_i, c_j) \succ 1 \leftrightarrow c_i > c_j$	More important	1
$\delta(c_i, c_j) \succ 1 \leftrightarrow c_i >> c_j$	Much more important	2
$\delta(c_i, c_j) \gg 1 \leftrightarrow c_i >>> c_j$	Absolutely more important	3

Let  $D$  be a set of decision-making agents,  $D = \{DM_1, DM_2, \dots, DM_k, \dots, DM_n\}$  that express their opinions on the relative importance of the criteria involved. These preferences give rise to the  $MDM_k$  preference matrix. The relationship between the two scales of the table allows the transformation of the matrix (1) into (2):

$MDM_k = [\delta(c_i, c_j)]$ , in a column vector  $[V_i]$ , where: (1)

$$V_i = \sum_{j=1}^m \delta(c_i, c_j) \quad (i = 1, \dots, m, \text{ and } k = 1, \dots, n) \quad (2)$$

After generating the vector  $V_i$ , its  $a_{ij}$  elements are normalized according to (3):

$$v = (a_{ij} - \min a_{ij}) / (\max a_{ij} - \min a_{ij}) \quad (3)$$

Giving rise to the  $DM_k$  preferences vector. If null values occur in this step, they are replaced by 1% of the second lowest value obtained. After all DM's carry out their evaluations, the normalized vectors are added, giving rise to the weight vector that expresses the importance of the criteria [10].

Step 2 - Ordinal classification ( $\Theta_{ij}$ ) of the performance of the alternatives:

In this step, each DM assigns the ratings related to the performance of the alternatives in each criterion (table 3), which are related to their rating ranges  $g_{(ij)}$ . After all "n" DMs perform their evaluations, the arithmetic mean  $\mu_{(ij)}$  of the classification ranges of the performances of the alternatives in each criterion is obtained.

**Table 3.** Ordinal ratings of performance of alternatives

Ordinal classification ( $\Theta_{ij}$ ) of the performance of alternative $i$ in criterion $j$	Classification range $g_{(ij)}$
Excellent (E)	1
Very Good (VG)	2
Good (G)	3
Medium (M)	4
Bad (B)	5
Very Bad (VB)	6
Poor (P)	7

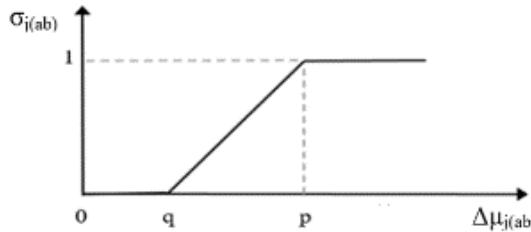
Step 3 – Obtaining the fractions of the criteria weights ( $\sigma_{j(ab)}$ ).

For each criterion "j", a parity comparison is made between the alternatives to verify the relative distance between the mean values of the classification ranges (4):

$$\Delta\mu_{i(ab)} = \mu_{(ia)} - \mu_{(ib)} \quad (4)$$

This value allows identifying in the preference modeling (figure 1 and table 4) the weight fraction of criterion "j", obtained by alternative "a" in relation to alternative "b" ( $\sigma_{j(ab)}$ ).

**Figure 1.** Preference function of a criterion with linear variation [11].



**Table 4.** Criteria preference modeling

Indiferença (I)	$\mu_{(ia)} - \mu_{(ib)} \leq 1$	$\rightarrow \sigma_{(ab)} \rightarrow 0$
Preferência Fraca (Q)	$1 < \mu_{(ia)} - \mu_{(ib)} \leq 3$	$\rightarrow \sigma_{(ab)} \rightarrow \frac{(A_{ij} - \min A_{ij})}{(\max A_{ij} - \min A_{ij})}$
Preferência Forte (P)	$3 < \mu_{(ia)} - \mu_{(ib)}$	$\rightarrow \sigma_{(ab)} \rightarrow 1$

Step 4 – Calculation of relative  $d_{ab}$  dominance.

Obtained by the weighted sum of the criteria weights ( $w_j$ ), with the corresponding fraction ( $\sigma_{j(ab)}$ ) verified in the preference modeling (5):

$$d_{ab} = \sum w_j \times \sigma_{j(ab)} \quad (5)$$

Step 5 - Calculation of Absolute  $D_{ab}$  Dominance and Outranking Rate  $\eta_{ab}$ :

The difference between the relative dominances  $d_{ab} - d_{ba}$  provides information on the absolute dominance  $D_{ab}$  between the alternatives (6).

$$D_{ab} = d_{ab} - d_{ba} \quad (6)$$

By dividing  $D_{ab}$ , by the sum of the weights, the percentage rate of absolute dominance between the alternatives is obtained.

Step 6 – Assembling the graph with the dominance relationships between the alternatives (Optional).

#### 2.4. Application of SAPEVO-M-NC

Tables 5 and 6 respectively present the values of the weights of the criteria, and the average performance of the alternatives after evaluation by  $DM_1$  and  $DM_2$ :

**Table 5.** Weights of the criteria

Criteria	DM <sub>1</sub>	DM <sub>2</sub>	Final weight
C <sub>1</sub> - Profitability	1	1	2
C <sub>2</sub> - Risk	0,667	0,714	1,381
C <sub>3</sub> - Liquidity	0,333	0,286	0,619
C <sub>4</sub> - Volatility	0,3	0,143	0,146
C <sub>5</sub> - Initial application	0,222	0,001	0,223

**Table 6.** Average performance of alternatives in the criteria

Investment	Profitability	Risk	Liquidity	Volatility	initial application
A <sub>1</sub> (FJPF)	6	1	1	1	1
A <sub>2</sub> (FJPR)	6,5	2	1,5	1,5	2
A <sub>3</sub> (FM)	1	4,5	5,5	2	5,5
A <sub>4</sub> (FA)	1	5,5	7	6,5	4
A <sub>5</sub> (FC)	2,5	5,5	1,5	4,5	1

Table 7 presents the performance of the alternatives, the dominance relationships, and the percentage of the composition of the investment portfolios.

**Table 7.** The dominance relationships and the percentage of the composition of the investment portfolios

$\Delta\mu_{(ab)}$		$\sigma_{(ab)}$						d <sub>ab</sub>				D <sub>ab</sub>	$\eta_{ab}$	% (d <sub>ab</sub> / $\sum d_{ab}$ )	
		C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>	C <sub>5</sub>	C <sub>6</sub>	C <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>	C <sub>10</sub>	Total			
A <sub>1</sub> - A <sub>2</sub>	0,50	1,00	0,50	0,50	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
A <sub>1</sub> - A <sub>3</sub>	-5,00	3,50	4,50	1,00	4,50	1,25	1,75	0,00	1,75	0,00	1,73	1,08	0,00	0,39	3,20
A <sub>1</sub> - A <sub>4</sub>	-5,00	4,50	6,00	5,50	3,00	1,75	2,50	2,25	1,00	0,00	2,42	1,55	0,33	0,22	4,52
A <sub>1</sub> - A <sub>5</sub>	-3,50	4,50	0,50	3,50	0,00	1,75		1,25	0,00	2,42	0,00	1,73	0,00	0,18	2,60
d <sub>ab</sub> - d <sub>ba</sub>														10,32	
A <sub>2</sub> - A <sub>1</sub>	-0,50	-1,00	-0,50	-0,50	-1,00					0,00	0,00	0,00	0,00	0,00	0,00
A <sub>2</sub> - A <sub>3</sub>	-3,50	2,50	4,00	0,50	3,50	0,75	1,50		1,25	0,00	1,04	0,93	0,00	0,28	2,24
A <sub>2</sub> - A <sub>4</sub>	-3,50	3,50	5,50	5,00	2,00	1,25	2,25	2,00	0,50	0,00	1,73	1,39	0,29	0,11	3,52
A <sub>2</sub> - A <sub>5</sub>	-4,00	3,50	0,00	3,00	-1,00	1,25		1,00	0,00	1,73	0,00	0,15	0,00	1,87	-1,13
d <sub>ab</sub> - d <sub>ba</sub>														7,63	
A <sub>3</sub> - A <sub>1</sub>	5,00	-3,50	-4,50	-1,00	-4,50	2,00				4,00	0,00	0,00	0,00	4,00	0,80
A <sub>3</sub> - A <sub>2</sub>	5,50	-2,50	-4,00	-0,50	-3,50	2,25				4,50	0,00	0,00	0,00	4,50	2,26
A <sub>3</sub> - A <sub>4</sub>	0,00	1,00	1,50	4,50	-1,50	0,00	0,25	1,75		0,00	0,00	0,15	0,26	0,00	0,41
A <sub>3</sub> - A <sub>5</sub>	1,50	1,00	-4,00	2,50	-4,50	0,25	0,00	0,75	0,50	0,00	0,00	0,11	0,00	0,61	-0,71
d <sub>ab</sub> - d <sub>ba</sub>														9,52	
A <sub>4</sub> - A <sub>1</sub>	5,00	-4,50	-6,00	-5,50	-3,00	2,00				4,00	0,00	0,00	0,00	4,00	-0,52
A <sub>4</sub> - A <sub>2</sub>	5,50	-3,50	-5,50	-5,00	-2,00	2,25				4,50	0,00	0,00	0,00	4,50	0,98
A <sub>4</sub> - A <sub>3</sub>	0,00	-1,00	-1,50	-4,50	1,50		0,25	0,00	0,00	0,00	0,00	0,06	0,06	-0,36	-0,17
A <sub>4</sub> - A <sub>5</sub>	1,50	0,00	-5,50	-2,00	-3,00	0,25			0,50	0,00	0,00	0,00	0,00	0,50	-1,19
d <sub>ab</sub> - d <sub>ba</sub>														9,06	
A <sub>5</sub> - A <sub>1</sub>	3,50	-4,50	-0,50	-3,50	0,00	1,25				2,50	0,00	0,00	0,00	2,50	-0,10
A <sub>5</sub> - A <sub>2</sub>	4,00	-3,50	0,00	-3,00	1,00	1,50				3,00	0,00	0,00	0,00	3,00	1,13
A <sub>5</sub> - A <sub>3</sub>	-1,50	-1,00	4,00	-2,50	4,50	1,50				0,93	0,00	0,39	1,32	0,71	0,33
A <sub>5</sub> - A <sub>4</sub>	-1,50	0,00	5,50	2,00	3,00	2,25	0,50	1,00	0,00	1,39	0,07	0,22	1,69	1,19	0,55
d <sub>ab</sub> - d <sub>ba</sub>														8,51	
$\sum d_{ab}$														45,04	

### 3. Analysis of results and Conclusion

The approach by the SAPEVO-M-NC method enabled a robust and efficient approach to solve the problem of structuring the investment portfolio in question, which considered, the evaluation of two analysts, the current scenario of the Brazilian economy and historical series of alternatives in the criteria in question, providing greater robustness and clarity to the investor in question. The financial contribution among the investments was distributed evenly, not showing on the part of the analysts a strong fear

that would incite them to direct investments towards low risk, abdicating greater profits, nor an excess of confidence in an accelerated economic recovery, to explore alternatives with greater risks and initial contributions to obtain greater profits.

Due to the relative ease of application, associated with a low cognitive effort on the part of the evaluators, it appears that this methodology can provide great gains, not only for the academic community, but also for society as a whole, presenting itself as a alternative tool to support multicriteria decision, of an ordinal, non-compensatory nature and with the possibility of supporting multiple decision makers.

To carry out a more in-depth evaluation, as a proposal for future work, it is proposed to use other methods of multicriteria decision aid in the case studied, making a comparison with the method used.

## References

- [1] de Melo CML, Silva GAS, Melo ARS, de Freitas AC. COVID-19 pandemic outbreak: The Brazilian reality from the first case to the collapse of health services. *An Acad Bras Cienc.* 2020;92(4):1–14.
- [2] Costa IP de A, Maeda SM do N, Teixeira LFH de S de B, Gomes CFS, dos Santos M. Choosing a hospital assistance ship to fight the Covid-19 pandemic. *Rev Saude Publica.* 2020;54, DOI: 10.11606/S1518-8787.2020054002792.
- [3] Yi C, Zhan Y, Zhang J, Zhao X. Ownership structure and OFDI by EMNES: the moderating effects of international experience and migrant networks. *Int J Emerg Mark.* 2021;(March).
- [4] Oliveira AS, Gomes CFS, Clarkson CT, Sanseverino AM, Barcelos MRS, Costa IPA, et al. Multiple Criteria Decision Making and Prospective Scenarios Model for Selection of Companies to Be Incubated. *Algorithms.* março de 2021;14(111), DOI: 10.3390/a14040111
- [5] Tenório FM, dos Santos M, Gomes CFS, Araujo J de C. Navy Warship Selection and Multicriteria Analysis: The THOR Method Supporting Decision Making. In: Springer Proceedings in Mathematics and Statistics. 2020. p. 27–39, DOI: 10.1007/978-3-030-56920-4\_3
- [6] Jardim RRAJ, Santos M, Neto ECDO, Da Silva ED, De Barros FCMM. Integration of the waterfall model with ISO/IEC/IEEE 29148:2018 for the development of military defense system. *IEEE Lat Am Trans.* 2020;18(12):2096–103, DOI: 10.1109/TLA.2020.9400437
- [7] De Barros MD, Salles CAL, Gomes CFS, Da Silva RA, Costa HG. Mapping of the scientific production on the ITIL application published in the national and international literature. Elsevier - Procedia Computer Science Procedia Computer Science 2015. p. 102–11, DOI: [10.1016/j.procs.2015.07.013](https://doi.org/10.1016/j.procs.2015.07.013)
- [8] Oliveira AO, Oliveira HLS, Gomes CFS, Ribeiro PCC. Quantitative analysis of RFID publications from 2006 to 2016. *Int J Inf Manage.* 2019;48:185–92, DOI: [10.1016/j.ijinfomgt.2019.02.001](https://doi.org/10.1016/j.ijinfomgt.2019.02.001)
- [9] Pereira F de C, Verocai HD, Cordeiro VR, Gomes CFS, Costa HG. Bibliometric Analysis of Information Systems Related to Innovation. *Procedia Comput Sci.* 2015;55:298–307. DOI: [10.1016/j.procs.2015.07.052](https://doi.org/10.1016/j.procs.2015.07.052)
- [10] Gomes CFS, Santos M Dos, Teixeira LFH de S de B, Sanseverino AM, Barcelos MRDS. Sapevo-m: A group multicriteria ordinal ranking method. *Pesqui Operacional.* 2020;40:1–23.
- [11] Maeda SM do N, Costa IP de A, Santos M dos., Gomes CFS. Economic and edaphoclimatic evaluation of Brazilian regions for African mahogany planting - an approach using the SAPEVO-M-NC ordinal method. In: Elsevier - Procedia Computer Science Procedia Computer Science. 2021.
- [12] Gomes CFS, Gomes LFAM. Princípios e métodos para a tomada de decisão: Enfoque multicritério. 6th ed. Atlas Humanas Didático; 2019. 360 p.

# Role of Big Data Analytics in Belt and Road Initiative (BRI): Multivariate Analysis with Gaussian Distribution of Data

Valliappan Raju <sup>a1</sup>, Wang Juan <sup>b</sup>, Sandeep Shrestha <sup>b</sup>, Arrunkumar Kalathinathan <sup>c</sup>, KK. Ramachandran <sup>d</sup>

<sup>a</sup> Limkokwing University of Creative Technology, Malaysia

<sup>b</sup> Qiqihar Medical University, China

<sup>c</sup> Research Associate, World Research Union, India

<sup>d</sup> GRD College of Arts & Science, India

**Abstract.** This manuscript focuses on the Belt and Road Initiative (BRI) of China, whereby the focus is on the engagement of big data analytics to comprehend logistics exertion. China is the trendsetter for revolutionary practices in trade, logistics, and technology. The recent progress the nation is thriving is on ‘One Belt One Road’ project whereby 65 countries are involved. It aims to connect continents and circulate smooth trade between them. This paper addresses the role of the database to identify the inter-model logistics in BRI. The merits of this project in the perspective of economic growth are measured through a quantitative study with 112 samples. Goal-setting theory is used to construct a conceptual framework for the research. Multivariate analysis is executed with SmartPLS 3.3.3 followed by an in-depth structural equation modeling. Normal distribution of data was given importance as in statistics the real-value of random variables whose distributions are not known, thus Gaussian distribution of data was used. Out of 6 Hypotheses, it is noted that five are significantly positive. Hypothesis testing is concluded based on p-value and t-statistics. The outcome of research suggests that big-data analytics is a major contributor in determining the significant model on logistics in Belt and Road Initiative.

**Keywords.** Big Data Analytics, Belt and Road Initiative, China, Economic Growth, Silk Road

## 1. Introduction

The Belt and Road Initiative (BRI) of China has emerged as one of the spectacular leaps in infrastructural development. Although several nations hold bilateral agreements to uplift their connectivity, China has gone far beyond expectation through its BRI project. The owners of the silk route prove to be sustainable in innovation and development [1]. It is to be noted that a study by CBER in 2019 forecasted that project BRI may boost global GDP (Gross Domestic Product) by \$7.1 trillion per annum by 2040 [2]. In the past

---

<sup>1</sup> Corresponding Author, Valliappan Raju, Limkokwing University of Creative Technology, Malaysia; Email: valliappan.raju@limkokwing.edu.my

two decades, the economy in China has undergone mammoth shifts positively. China has agreed to disseminate its funding up to USD 1.25 trillion by 2025, which shall be the highest ever FDI that has happened from the USA. The main objective was to build a singular and competent market to set up a trading place where both domestic and international commerce takes place through unity, integrating technology and cultural exchange [3]. To achieve this, it is important to understand the nuances of this project by comprehending the role of big data [4]. After completion of this project, one can evaluate the cost-opportunity of shipping, consigning, and traveling [5]. If this logistics is done effectively, then it adds more value. Through this research paper, the authors enumerate that by using Big data analytics, prominent logistics can adhere.

### *1.1 Overview of Big Data Analytics*

Big Data Analytics comprises two distinct verbs, namely, Big Data and Analytics. Big data represents a lot of data. However, big data is not defined only in terms of volume (storage capacity) but also variety (different types of data) and velocity (the speed at which data is created). Big data can be acquired from multiple sources – web log files from the web, text data from surveys, geospatial data from mobile phones, social media are some examples [6]. Hence, data can sometimes be collected specifically for a certain form of analytics [7]. Put together, Big Data Analytics is an extremely powerful tool to understand and predict.

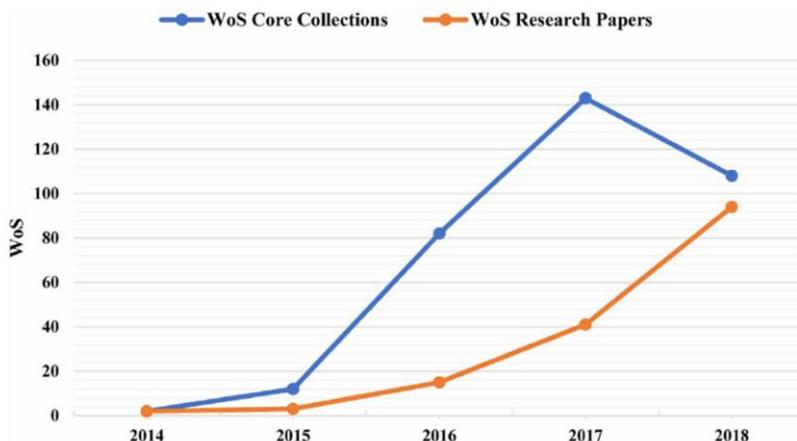
### *1.2 Big Data Analytics in Logistics*

Gone are the days when Logistics used to be substandard due to data imprecision or usage of EPR systems for maintaining warehouses or unorganized inventories or limited communication with consumers or unpredictable traffic patterns and weather conditions or unpredictable market behavior [8]. Big Data Analytics has made significant benefits in Logistics. Analytics of this data leads to a better understanding of purchase patterns of various consumers, the latest demands in the market, and maintenance cycles [9].

## **2. Literature Review**

The core purpose of a literature review is to identify scholarly inputs in the respective area of research and to provide immense knowledge to readers about the respective research [10]. While reviewing, strengths and weaknesses can be comprehended. A meticulous review of the literature was done on Belt and Road Initiative (BRI) and Big Data Analytics (BDA). Coherently both these concepts are new and limited literature was found [9]. However, several scholars have aligned their views that new technologies can build economies with rapid growth. It has to be understood that Machine Learning is part of Big Data, thus Machine Learning plays a dominant role [4]. In the context of BRI, the codes for machine learning are to be used during the implementation phase. With this seed of thought, literature from 2014 was considered, the reason being, project BRI was officially initiated in the year 2013 [1].

There are 347 research papers about Belt and Road Initiative (BRI) as of October 2018 particularly which are indexed with Web of Science core collection journals [3]



**Figure 1- BRI on WoS Database (Source: Shahriar, 2018)**

Figure 1 illustrates the bibliographic analysis between Web of Science core collections and research papers. Furthermore, below Table 1 represents the details of 32 research manuscripts that dealt with about Belt and Road Initiative along with their findings

The key five points from the CEBR study pointed out that,

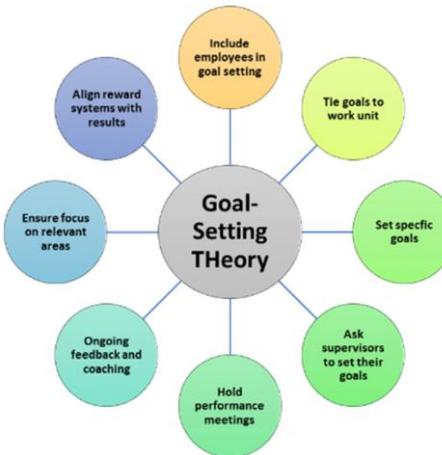
- 1) By this project, BRI China may continue to lead the world economy in terms of production, infrastructure, employment rate etcetera. With this, China has a high chance to lead the World sustainably for the next few decades [11], [12]
- 2) The advantages of project BRI are meritorious because 70 countries have agreed to be part of it, and around 56 countries forecast that their GDP will be escalated to USD 10 billion and beyond by 2040.
- 3) The impact of this project can trigger economical shifts in Russia, India, Japan, Korea, Indonesia, Netherlands, and United Kingdom [13].
- 4) The United States of America could be a potential beneficiary from this project BRI though it has no direct involvement. Thus by 2040, the USA also has the potential to get benefitted from project BRI [14]
- 5) Planet Earth is supposed to be transformed by BRI. This makes the clear notion that project BRI has the meritocracy to change the landscape of business in the World [15]

## 2.1 Underpinning Theory

Edwin A. Locke constructed Goal Setting theory in the year 1986. Goal Setting Theory has been considered in this research as underpinning theory as one of the main components to distinguish between Belt and Road Initiative and Big Data Analytics [16].

### 2.1.1 Rationale of Underpinning Theory

The by-product of this research is on project performance, that is, by identifying the role of big data analytics in BRI project, the ultimate aim is to successfully complete the project. Big Data is a means [17]. Therefore, it would be right to presume that ‘project

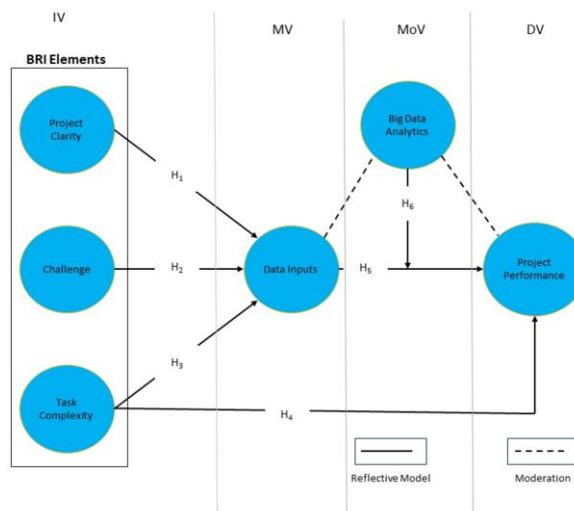


**Figure 2-** Goal Setting Theory Model

performance' is the dependent variable for this research. To measure and successfully complete a project, the short-term and long-term goals has to be set. Thus, this theory has been adopted in this respective research.

## 2.2 Conceptual Framework

Based on Goal Setting Theory, a conceptual framework has been developed for respective research whereby the core purpose is to relate big data analytics in Belt and Road initiative (BRI). Here, BRI is considered as the principal (main goal), and big data is considered as an agent (processor). In this connotation, the variables were retrieved from literature and the above-mentioned theories.



**Figure 3-** Conceptual Framework

### 3. Research Methodology

The independent variables in the above conceptual framework were randomly reviewed and enforced to have a normal distribution of data in data collection [18]. Gaussian distribution of data means to be normally distributed, that is, to normally deviate without hitting extremes [1]. The standard deviation to be within +1 to -1 as the curve to be in bell shape. This can be known only after data is collected.

The sample size can be arrived based on following formula,

$$n = N * \frac{X}{X+N-1}$$

Where  $N$  = population size;  $X$  = target population

However, for this research, data were collected from 112 samples which were based on Krejcie & Morgan's (1970) sample-size calculation method. These 112 respondents are from IT, Supply chain management, Chinese diplomats, and economists. Thus, it is a stratified random sampling that consists of four different strata namely IT industry, personnel from the logistics sector, diplomats from the Chinese embassy in ASEAN, and lastly from economists who possess technical understanding towards research analysis [19].

### 4. Data Analysis

The analysis is done to determine the project performance of BRI using Big Data Analytics as moderator and data inputs as mediator [20]. While the motive is to comprehend a prominent model to spur trade across the globe, inter-model transportation was one familiar strategy [21]. Adding to this, if big data analytics are done then the outcome is larger than inter-model transportation. To testify this, 112 respondents were surveyed, and data were collected.

#### 4.1 Measurement Model

The measurement model is also called an outer model, whereby the indicators or items are analyzed with their significance with constructs. Reliability, validity, R square adjusted, multicollinearity is analyzed. Reliability is about consistency and validity is about the accuracy of items. Table 1 illustrates the reliability and validity

**Table 1** – Construct Reliability and Convergent Validity

	Cronbach's Alpha	rho_A	Composite Reliability	Average Variance Extracted (AVE)
<b>BigData_Analytics</b>	<b>0.850</b>	<b>0.864</b>	<b>0.912</b>	<b>0.777</b>
<b>Challenge</b>	<b>0.915</b>	<b>0.938</b>	<b>0.941</b>	<b>0.766</b>
<b>Data_Inputs</b>	<b>0.830</b>	<b>0.841</b>	<b>0.899</b>	<b>0.748</b>
<b>MD-DI*BD*PP</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
<b>Project_Clarify</b>	<b>0.858</b>	<b>0.914</b>	<b>0.908</b>	<b>0.683</b>
<b>Project_Performance</b>	<b>0.813</b>	<b>0.842</b>	<b>0.872</b>	<b>0.584</b>
<b>Task_complexity</b>	<b>0.886</b>	<b>0.911</b>	<b>0.914</b>	<b>0.682</b>

(Source: Generated from SmartPLS 3.3.3)

#### 4.2. Parameters for Measurement Model

The threshold for reliability is <0.7 and the threshold for convergent validity is <0.5. In Table 3, the reliability is identified by both Cronbach's Alpha and Composite Reliability. Cronbach's Alpha is calculated for vetting Confirmatory[22]. Factor Analysis (CFA) whereas Composite Reliability is calculated for vetting Exploratory Factor Analysis (EFA). Convergent Validity is calculated through AVE. According to [23] the threshold for reliability (represented by Cronbach's alpha) is it should be more than 0.7. By this, all variables are reliable, as the least value is 0.813 (for Project Performance). Similarly, composite reliability needs to be greater than 0.7, which is also complied with. Average Variance Extracted (AVE) is the convergent validity of items or indicators. The threshold for AVE has to be greater than 0.5 [24] Above table reflects a similar outcome, so the validity of the instrument is high. The square root of the AVE value will be the same as the respective construct's value in discriminant validity. Discriminant validity represents the accuracy of items in the research instrument. Following Table 2 highlights of Discriminant Validity

**Table 2 – Discriminant Validity**

Fornell-Larcker Criterion								
	BigData Analytics	Challenge	Data_Inputs	MD-DI*BD *PP	Project_Clarity	Project_Performance	Task_complexity	
<b>BigData_Analytics</b>	0.881							
<b>Challenge</b>	0.801	0.875						
<b>Data_Inputs</b>	0.879	0.928	0.865					
<b>MD-DI*BD*PP</b>	-0.147	-0.447	-0.423	1.000				
<b>Project_Clarity</b>	0.905	0.940	0.971	-0.338	0.826			
<b>Project_Performance</b>	0.945	0.877	0.931	-0.219	0.958	0.764		
<b>Task_complexity</b>	0.913	0.846	0.930	-0.183	0.945	0.954	0.826	

(Source: Generated from SmartPLS 3.3.3)

The value of construction in each column has to be more than the value of the rest for discriminant validity. There are certain values lesser, namely, the project performance has got the value of 0.764 whereas the other value is 0.954. This proves that the validity is weak, however, for other constructs, it's within the limit. VIF (Variance Inflation Factor) measures the multicollinearity of indicators of constructs. According to [24] the VIF in the outer range should not be more than 10 whereas according to [25] 5 is the threshold.

**Table 3 – Multicollinearity (VIF)**

	VIF
<b>BD1</b>	<b>1.414</b>
<b>DI1</b>	<b>1.558</b>
<b>Data_Inputs * BigData_Analytics</b>	<b>1.000</b>
<b>PC1</b>	<b>9.127</b>
<b>PP1</b>	<b>2.571</b>
<b>TC1</b>	<b>7.542</b>

(Source: Generated from SmartPLS 3.3.3)

In this analysis, only one item in each construct is considered. Coherently convincing both these scholarly thresholds, the respective research has less than 5 for all constructs except for Project Complexity. This proves that the VIF is at par with the threshold. Above Table 3 highlights the VIF values.

#### 4.3 Structural Modeling

Structural Modeling is also called an inner model. Structural modeling in Multivariate analysis comprises mediation analysis through PLS-SEM (Partial Least Square-Structural Equation Modeling), moderation analysis [26]. Data Input is the mediator and it is hypothesized between the independent variable and dependent variable. Results of indirect effects in a model can yield the potency of the mediating effect. Following table 4 highlights the values on indirect effects

**Table 4** – Indirect Effects (Mediating Effects)

	Original Sample (O)	Sample Mean (M)	Standard Deviation (STDEV)	T Statistics ( O/STDEV )	P Values
Task_complexity -> Data_Inputs -> Project_Performance	0.112	0.116	0.052	2.167	<b>0.031</b>
Challenge -> Data_Inputs -> Project_Performance	0.105	0.110	0.047	2.221	<b>0.027</b>
Project_Clarity -> Data_Inputs -> Project_Performance	0.328	0.328	0.083	3.958	<b>0.000</b>

(Source: Generated from SmartPLS 3.3.3)

From Table 4 it is evident that Data Input mediates between the constructs efficiently. Based on the p-value, which is lesser than 0.05 it can be understood that all three independent variables require a mediator to achieve project performance (dependent variable).

#### 4.4 Hypothesis Testing

The vital section in research is testing the claims or hunches or assumptions. Hypothesis testing is done for this respective research as this is more of quantitative research. A familiar or renowned technique to interpret the hypothesis testing is based on a p-value score. The P-value score is based on T-statistics. These both go hand-in-hand. P-value (probability value) has to be lesser than 0.05, which is 95% has positivity and 5% is doubtful. Thus, as long as 5% is in the doubtful scenario, the model is accepted.

### 5. Conclusion

Belt and Road Initiative (BRI) is a visionary project initiated by President Xi Jinping of China. After several challenges and constraints, this project has come to the construction stage in 2020. To amplify this project, there are strategies in inter-model logistics. However, using data from past trades in silk routes, there can be an optimized solution than inter-model [27]. If Big Data Analytics is used, this model will be more successful. However, this was merely the author's opinion before drafting this research manuscript.

After indulging in a data collection with 112 respondents who are involved with BRI, the data were analyzed and found that respondents abide that big data analytics will amplify the economic bottom line of the project and also enhance the performance. Big Data Analytics was not tested as an independent variable, whereas, it was placed as a moderator. Through moderation analysis, it was found that the p-value was 0.011 and t-statistics was 2.560, which is considered highly significant.

## Reference

- [1] V. Raju, "Implementing Flexible Systems in Doctoral Viva Defense Through Virtual Mechanism," *Glob. J. Flex. Syst. Manag.*, vol. 22, no. 2, pp. 127–139, Jun. 2021, doi: 10.1007/s40171-021-00264-y.
- [2] C. Policy, "Belt and Road Initiative to boost world GDP by over \$7 trillion per annum by 2040 - CEBR." <https://cebr.com/reports/belt-and-road-initiative-to-boost-world-gdp-by-over-7-trillion-per-annum-by-2040/> (accessed Aug. 09, 2021).
- [3] S. Shahriar, *Literature Survey on the "Belt and Road" Initiative: A Bibliometric Analysis*. Northwest A&F University, China, 2019.
- [4] A. Petraška, K. Čižiūnienė, A. Jarašūnienė, P. Maruschak, and O. Prentkovskis, "Algorithm for the assessment of heavyweight and oversize cargo transportation routes," *Vilnius Gedim. Tech. Univ.*, vol. 18, no. 6, pp. 1098–1114, Nov. 2017, doi: 10.3846/16111699.2017.1334229.
- [5] V. Raju and S. P. Poh, "Sustainability in Performance Management through Supply Chain Management," *Int. J. Supply Chain Manag.*, vol. 8, no. 2, pp. 1085–1089, Apr. 2019, Accessed: Jul. 08, 2021. [Online]. Available: <http://www.ojs.excelingtech.co.uk/index.php/IJSCM/article/view/2866>.
- [6] R. Perez-Vega, V. Kaartemo, C. R. Lages, N. Borghesi Razavi, and J. Männistö, "Reshaping the contexts of online customer engagement behavior via artificial intelligence: A conceptual framework," *J. Bus. Res.*, Nov. 2020, doi: 10.1016/j.jbusres.2020.11.002.
- [7] F. Visconti and J. M. de Paz, "Electrical Conductivity Measurements in Agriculture: The Assessment of Soil Salinity," in *New Trends and Developments in Metrology*, InTech, 2016.
- [8] G. Rampersad, "Robot will take your job: Innovation for an era of artificial intelligence," *J. Bus. Res.*, vol. 116, pp. 68–74, Aug. 2020, doi: 10.1016/j.jbusres.2020.05.019.
- [9] V. Raju and S. Poh Phung, "Resilient Motive for Reviewing Literatures in Business Research: Exploring Process of Research Design," Apr. 2021. Accessed: Jul. 05, 2021. [Online]. Available: <http://annalsofrscb.ro>.
- [10] J. E. Adaletey, V. Raju, and S. P. Phung, "Role of stakeholder in revenue mobilization to alleviate poverty in Ghana using E-governance mechanisms," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 2S, 2018.
- [11] B. Kar, "Service Quality and SERVQUAL Model: A Reappraisal," *Amity J. Oper. Manag. ADMAA*, vol. 1, no. 2, pp. 52–64, 2016.
- [12] M. J. Alexander, E. Jaakkola, and L. D. Hollebeek, "Zooming out: actor engagement beyond the dyadic," *J. Serv. Manag.*, vol. 29, no. 3, pp. 333–351, 2018, doi: 10.1108/JOSM-08-2016-0237.
- [13] M. Lee and L. J. Kray, "A gender gap in managerial span of control: Implications for the gender pay gap," *Organ. Behav. Hum. Decis. Process.*, vol. 167, pp. 1–17, Nov. 2021, doi: 10.1016/j.obhdp.2021.06.001.
- [14] R. Flepp, P. Meier, and E. Franck, "The effect of paper outcomes versus realized outcomes on subsequent risk-taking: Field evidence from casino gambling," *Organ. Behav. Hum. Decis. Process.*, vol. 165, pp. 45–55, Jul. 2021, doi: 10.1016/j.obhdp.2021.04.003.
- [15] X. Guan, M. Liu, and Y. Meng, "A comprehensive ecological compensation indicator based on pollution damage – protection bidirectional model for river basin," *Ecol. Indic.*, vol. 126, p. 107708, Jul. 2021, doi: 10.1016/j.ecolind.2021.107708.
- [16] K. Plys, "Theorizing capitalist imperialism for an anti-imperialist praxis towards a rodneyan world-systems analysis," *J. World-Systems Res.*, 2021, doi: 10.5195/jwsr.2021.1022.
- [17] V. Raju and S. P. Phung, "Economic dimensions of blockchain technology: In the context of extention of cryptocurrencies," *Int. J. Psychosoc. Rehabil.*, vol. 24, no. 2, pp. 29–39, 2020, doi: 10.37200/IJPR/V24I2/PR200307.
- [18] M. R. Jalilvand, J. Khazaei Pool, H. Balouei Jamkhaneh, and R. A. Tabaeian, "Total quality management, corporate social responsibility and entrepreneurial orientation in the hotel industry," *Soc. Responsib. J.*, vol. 14, no. 3, pp. 601–618, 2018, doi: 10.1108/SRJ-04-2017-0068.

- [19] N. Bin Chang, K. Bai, and C. F. Chen, "Integrating multisensor satellite data merging and image reconstruction in support of machine learning for better water quality management," *J. Environ. Manage.*, vol. 201, pp. 227–240, 2017, doi: 10.1016/j.jenvman.2017.06.045.
- [20] S. U. A. Rana and V. Raju, "Factors influencing glass ceiling focus on women administration in higher education in Malaysia," *Int. J. Eng. Adv. Technol.*, 2019, doi: 10.35940/ijeat.F1220.0986S319.
- [21] R. W. and F. D. D. Zhang, M , Griffeth, "Corporate social responsibility, continuous process improvement orientation, organizational commitment and turnover intentions," *Int. J. Qual. Reliab. Manag.*, vol. 25, no. 2010, 2011, doi: 10.1108/JHOM-09-2016-0165.
- [22] J. A. G. J. T. Quin, "Comparison of Transformational Leadership Practices: Implications for School Districts and Principal Preparation.," *J. Leadersh. Educ.*, vol. 14, no. 3, pp. 71–85, 2015.
- [23] J. F. Hair Jr, G. T. M. Hult, C. Ringle, and M. Sarstedt, *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage Publications, 2017.
- [24] J. F. Hair, "Successful Strategies for Teaching Multivariate Statistics," *Proc. 7th Int. Conf. ...*, pp. 1–5, 2006.
- [25] C. Ringle, S. Wende, and J. Becker, "Ringle, Christian M., Wende, Sven, & Becker, Jan-Michael. (2015). SmartPLS 3. Bönningstedt: SmartPLS. Retrieved from <http://www.smartpls.com>," *Retrieved from*, 2015.
- [26] K. A. Law, A. Bhaumik, P. Sun, and V. Raju, "Factors determining the relationship between superiors and their subordinates: Evaluating the trust factor in Chinese organizations," *Int. J. Psychosoc. Rehabil.*, vol. 24, no. 4, 2020, doi: 10.37200/IJPR/V24I4/PR201082.
- [27] J. Lu, L. Ren, C. Zhang, D. Rong, R. R. Ahmed, and J. Streimikis, "Modified Carroll's pyramid of corporate social responsibility to enhance organizational performance of SMEs industry," *J. Clean. Prod.*, vol. 271, p. 122456, 2020, doi: 10.1016/j.jclepro.2020.122456.

# Research on the Impact of Timing Financing on M&A Performance: The Intermediary Effect Based on Financing Efficiency

SHANG Qianqian<sup>1</sup>

*School of Economic, Shanghai University, China*

**Abstract:** Timing financing theory holds that when listed companies are overvalued, they will increase equity financing. Different from previous studies, this study explores the mechanism of timing financing theory in China from the perspective of financing efficiency. Through the empirical study of M & A cases of Listed Companies in China from 2007 to 2018, it is found that the timing financing behavior that increasing equity financing when overvalued reduces the positive impact of equity financing on M & A performance. Further research found that the mediating effect of the financing efficiency exists, that is, the timing financing behavior does not increase listed companies' financing efficiency, so it can not help listed companies create value through M & A.

**Keywords:** Timing Financing Theory; equity financing; M&A performance; financing efficiency

## 1. Introduction

M & A is one of the most important investment methods of listed companies (Li, 2020)[1], whether financing can be obtained at a lower cost and whether sufficient funds can be obtained directly affect M & A performance. At present, the mispricing of the stock market caused by inefficient capital market has attracted more and more attention from scholars. A large number of empirical evidences show that the window period when the stock price of listed companies deviates from the fundamentals has become one of the important considerations for companies to make investment and financing decisions. Stein's (1996) timing financing theory holds that companies can make use of window opportunities caused by market irrationality to arrange financing decisions rationally[2]. Shleifer & Vishny (2003) extended the influence of timing financing theory from the perspective of M & A[3]. They believed that companies whose stock prices were overvalued would purchase assets by issuing shares or exchange shares to reducing financing costs and achieving limited arbitrage.

What kind of impact will the overestimation of M&A companies bring to M&A performance? Scholars have not reached a consistent conclusion on this question. Moreover, there are few researches on the influence mechanism of stock mispricing on M&A performance( Luo & Yang,2019)[4]. This paper explains the impact of stock

---

<sup>1</sup> Corresponding Author: SHANG Qianqian, School of Economic, Shanghai University, China; E-mail: shangqian@shu.edu.cn.

mispricing on M & A performance from the perspective of financing efficiency. Many studies in China have confirmed that equity financing can increase M & A performance. However, the paper found that overestimation of listed companies will reduce the positive impact of equity financing on M & A performance. This conclusion provides evidence for the timing theory of China's capital market, and enriches the research of equity financing on M&A performance.

In terms of financing efficiency, the degree of marketization and property rights system in developed capital market is relatively sound, so there are few studies on financing efficiency( Shen, 2017; Yin et al., 2018)[5][6]. But due to the constraints of institutional environment, corporate governance level and financing in China's capital market, the financing efficiency is generally low (Zhang et al., 2017)[7]. Therefore, different from foreign research, this study explains the impact of corporate mispricing on M&A performance from the perspective of financing efficiency. It is found that the reason for overestimation reduces the positive impact of equity financing on M&A performance is that overestimation reduces financing efficiency, which in turn reduces M&A performance. It means the mediating mechanism of financing efficiency on the influence of overvaluation on M&A performance is established. The research on the impact of financing efficiency on timing financing is more in line with the background of China's capital market, and it also expands the mechanism research scope of the impact of mispricing on M&A performance.

At present, the single financing method is the main problem of M & A financing in China, mainly self-owned funds. In the future, with the formal implementation of the new securities law in China's capital market and the gradual relaxation of equity financing channels, the proportion of equity financing will increase, and the market timing behavior will be more extensive and common. Therefore, studying how to reasonably use capital market valuation deviations, reduce financing costs, improve financing efficiency, and achieve better M&A performance is of great practical significance.

## 2. Theoretical analysis and research hypothesis

Since 1990s, the most important research direction of financing has been the impact of market mispricing on corporate financing decisions (Huang & Ritter, 2005)[8]. According to behavioral corporate finance theory, inefficient market condition is an important factor in managers' financing decisions, that is, companies will choose financing strategies based on the degree of deviation between stock prices and fundamentals.

The first influential theory that the capital market affects corporate financing choices is Stein's (1996) timing financing theory. By constructing a rational capital budget analysis framework under the irrational capital market, Stein (1996) found that mispricing in the stock market will affect the companies' equity financing costs, so companies can choose financing methods based on market valuation deviations<sup>[2]</sup>. When the stock price is higher than fundamentals, companies will choose equity financing. Otherwise, companies will choose debt financing or repurchase shares. Then the market timing financing has been confirmed, Alti (2006), Hovakimian (2006) and Kayhan & Titman (2007) all considered that companies have market timing behaviors when issuing securities[9][10][11]. Vagenas (2020) believes that companies with overvalued stock prices prefer to use share-based payments, and ultimately achieve the

goal of arbitrage from high-valued stocks[12].

Some scholars in China have found that market mispricing does play an important role in equity payment(Sun & Wei,2021)[13]. Specific to China's capital market, although there are many financing options available, the financing methods that meet the company's financing needs are very limited. Companies generally have a funding gap in China. That's because companies have low efficiency in the allocation of credit funds, and there are many restrictions on equity financing. This study believes that although timing financing can help listed companies to alleviate the financing constraints, but due to the imperfect internal governance of most listed companies in China, such as the absence of major shareholders or the high ownership concentration, the management lacks strategic M & A motivation, and the timing financing often has a certain degree of blindness. Therefore, the study conjectures that the timing financing will cause companies to deviate from the optimal capital structure and reduce the financing efficiency of listed companies, so it will not help increase M & A performance. So we assume that:

**H1:** Timing financing will not help listed companies increase M&A performance.

**H2:** Timing financing affects M&A performance through reducing financing efficiency.

### 3. Research Design

#### 3.1 Sample selection and data source

The M&A samples and their financial data come from CSMAR and Wind. There are 5,616 A-share M&A samples in 2004-2018. Because the impact of M & A lasts for 3 years, so only retain the case with the largest scale of M & A within 4 years; Considering that split share structure basically ended in 2006, so delete M & A cases before 2006; Delete financial institutions; Delete debt restructuring, asset replacement and asset divestiture cases; Delete ST and \* ST companies.

#### 3.2 Variable definition

##### 3.2.1 Company mispricing

In this study, we use the method of Rhodes-Kropf et al. (2005)[14] to decompose  $M/B$ :

$$M/B = (M/V) * (V/B) \quad (1)$$

Taking the logarithm on both sides, let  $\log M = m$ ,  $\log V = v$ ,  $\log B = b$ , the equation is converted into:

$$m - b = (m - v) + (v - b) \quad (2)$$

( $m-v$ ) is the deviation between market value and intrinsic value, that is the company's valuation deviation; ( $v-b$ ) is the deviation between intrinsic value and book value, that is the internal growth opportunity of the company. According to the literature, the deviation between market value and intrinsic value may come from the company level or the industry level. Equation (2) is transformed into:

$$m_{i,t} - b_{i,t} = [m_{i,t} - v(\theta_{i,t}; \alpha_{j,t})] + [v(\theta_{i,t}; \alpha_{j,t}) - v(\theta_{i,t}; \alpha_j)] + [v(\theta_{i,t}; \alpha_j) - b_{i,t}] \quad (3)$$

$v(\theta_{i,t}; \alpha_{j,t})$  represents the estimated value of the firm i in year t based on the estimation coefficient  $\alpha_{j,t}$  of the industry j in year t;  $v(\theta_{i,t}; \alpha_j)$  represents the estimated long-term value of the firm based on the estimation coefficient  $\alpha_j$  of the long industry j.

Therefore,  $[m_{i,t} - v(\theta_{i,t}; \alpha_{j,t})]$  represents the mispricing at company level;  $[v(\theta_{i,t}; \alpha_{j,t}) - v(\theta_{i,t}; \alpha_j)]$  represents the mispricing at industry level;  $[v(\theta_{i,t}; \alpha_j) - b_{i,t}]$  represents the firm's growth opportunities.

This study selects the mispricing at company level, that is  $[m_{i,t} - v(\theta_{i,t}; \alpha_{j,t})]$ . So we need calculate the estimated value of the firm i in year t  $v(\theta_{i,t}; \alpha_{j,t})$ , regress the following equation by year and industry:

$$m_{it} = \alpha_{0jt} + \beta_{1jt} b_{it} + \beta_{2jt} \ln(NI)_{it}^+ + \beta_{3jt} I_{(<0)} \ln(NI)_{it}^+ + \beta_{4jt} LEV_{it} + \beta_{5jt} LEV_{it}^2 + \varepsilon_{it} \quad (4)$$

$m_{it}$  is the market value of firm i at the end of year t, expressed by the logarithm of the sum of the equity market value and the bond market value;  $b_{it}$  is the logarithm of the total book assets of firm i at the end of year t;  $(NI)_{it}^+$  is the net profit after non-recurring profit and loss deduction where NI is positive only;  $I_{(<0)}$  is dummy variable, if NI is negative,  $I_{(<0)}$  is 1; otherwise,  $I_{(<0)}$  is 0.  $LEV_{it}$  is the financial leverage of company i in year t, which is (total assets - shareholders' equity) / total assets.

Regress formula (4) by industry and year, then get the estimated value of company i in year t:

$$\widehat{m}_{it} = \widehat{\alpha_{0jt}} + \widehat{\beta_{1jt}} b_{it} + \widehat{\beta_{2jt}} \ln(NI)_{it}^+ + \widehat{\beta_{3jt}} I_{(<0)} \ln(NI)_{it}^+ + \widehat{\beta_{4jt}} LEV_{it} + \widehat{\beta_{5jt}} LEV_{it}^2 \quad (5)$$

The estimated value estimated by formula 5 only includes company-specific factors, not industry valuation deviation, so the residual term  $\varepsilon$  of the model is the valuation deviation caused by company factors (Misping<sub>it</sub>):

$$\text{Misping}_{it} = m_{it} - \widehat{m}_{it} \quad (6)$$

### 3.2.2 M & A Performance

This paper uses EVA (Economic Value Added) to measure M & A Performance. Compared with the traditional financial indicator, it considers the capital cost, which can more measure whether an investment truly creates value for the enterprise (Zuo & Lv, 2014; Zhao & Li, 2017) accurately[15][16]. This paper chooses EVA<sub>1</sub>, EVA<sub>2</sub>, EVA<sub>3</sub> to represent the M & A performance within three years after M & A.

Learn from the methods mentioned in *Market Value Management Theory* by Shi and Liu (2008)[17]:

$$\text{EVA} = [\text{ROIC} - \text{WACC}] * \text{IC} \quad (7)$$

ROIC means Return on Invested Capital; WACC means Weighted Average Cost of Capital; IC means Investment Cost. Both ROIC and IC can be obtained directly in the financial statements. WACC can be obtained through Formula (8).

$$WACC = K_D * (1-T) * D / (D+E) + K_E * E / (D+E) \quad (8)$$

$K_D$  means debt interest rate, this paper uses the 3-5 year bank loan interest rate; T is the income tax rate at the end of each year;  $K_E$  means the cost of equity that uses CAPM(Formula (9)) to measure.

$$K_E = R_f + \beta(R_m - R_f) \quad (9)$$

$R_f$  means risk-free rate, this paper uses one year bank deposit interest rate in the year of M&A.  $\beta$ means risk coefficient,  $R_m$  means market return rate, that Shanghai Composite Index at the end of each year in this paper.

### 3.2.3 Financing Efficiency

Based on the research of financing efficiency factors by Wang et al. (2016)[18], this study selects DEA-Malmquist index to calculate the intertemporal financing efficiency change. The financing efficiency in the year of M & A is defined as the dummy variable Fiance0. Take 1 for the total factor production change (Tfpch) is greater than 1 in the year of M & A, it indicates that the financing efficiency has increased compared with the year before M&A, otherwise 0.

### 3.2.4 Control Variables

The study also controls other related variables such as YAER and INDUSTRY. The definitions and explanations of all variables are as follows:

**Table 1.** Description of main variables

Type	Name	Operational Definition
Dependent Variables	M & A Performance	$Eva1_i, Eva2_i, Eva3_i$ mean in the first, second and third years after M & A respectively, take logarithm;
Independent Variables	Mispricing	Define continuous variable to measure the extent of mispricing: $Conmisp = m_{it} - \widehat{m}_{it}$ ( $m_{it} - \widehat{m}_{it}$ , See formula 6 for details); Define dummy variable Highmisp, take 1for $m_{it} - \widehat{m}_{it} > 0$ , otherwise 0; Take data from the year before M & A;
Mediating Variables	Financing Efficiency	Selects DEA-Malmquist index to calculate the intertemporal efficiency change of M & A company's financing. Take 1 for the total factor production change (Tfpch) is greater than 1 in the year of M & A, otherwise 0.
Control Variables	Equity Financing	Take 1 for equity financing, otherwise 0;
	Property	Take 1 for state-owned enterprises in the year before M & A, otherwise 0;
	Listing Board	Take 1 for listing on the main board, otherwise 0;
	Profit	Average main business profit margin in three Years before M & A;
	Capital Structure	Average asset-liability ratio in three Years before M & A;
	Operating Capacity	Average total asset turnover rate in three Years before M & A;

---

Company Growth	Average revenue growth rate in three Years before M & A;
Corporate Governance	Ratio of the number of independent directors to the total number of board members in the year before M & A
Acquisition Size	Ratio of M & A Value to Market Value at the end of the year before M & A;
Rela	Take 1 for the acquirer and the acquiree are related parties, otherwise 0;

---

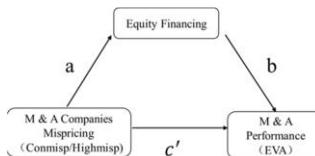
### 3.3 Model construction

$$\text{Eva1}_i/\text{Eva2}_i/\text{Eva3}_i = \beta_0 + \beta_1 \text{Equity Financing}_i + \beta_2 \text{Highmisp}_i * \text{Equity Financing}_i + \beta_3 \text{Controls}_i + \varepsilon_{it} \quad (10)$$

$$\text{Eva1}_i = \beta_0 + \beta_1 \text{Highmisp}_i + \beta_2 \text{Financing Efficiency}_i + \beta_3 \text{Controls}_i + \varepsilon_{it} \quad (11)$$

For H1, we choose OLS model to regress the formula (10).

For H2, we select  $\text{Eva1}_i$  as the dependent variable and select formula (10) and (11) to test the intermediary mechanism of financing efficiency. We refer to the method of Wen and Ye (2014) to test the mediating effect, the mediating effect of M&A financing efficiency is shown in Figure 1[19].



**Figure 1.** Test of the mediating effect of M&A Financing Efficiency

Path a indicates the impact of the independent variable on the mediating variable; path b indicates the impact of the mediating variable on the dependent variable; path c' indicates the impact of the independent variable on the dependent variable when consider the mediating variable. If both path a and path b are significant, the mediation effect is significant. If at least one of path a and path b is not significant, use the Bootstrap method to check whether the product of the coefficients is significant.

## 4. Empirical results and analysis

### 4.1 Descriptive statistics analysis

The results of descriptive statistics are reported in Table 2. Through descriptive statistics, it is found that EVA after M & A is increasing year by year, which shows that M & A in China has created value for companies. Highmisp, which is dummy variable that represents the company level mispricing, its mean is 0.79, which shows that listed companies in China are overvalued. The median of financing efficiency was 0, indicating that more than half of the listed companies' financing efficiency decrease in the year of M&A.

**Table 2.** Descriptive statistics

<b>Variables</b>	<b>N</b>	<b>mean</b>	<b>sd</b>	<b>min</b>	<b>p50</b>	<b>max</b>
Eva1	1948	23.81	1.350	17.90	23.76	29.87
Eva2	1588	23.87	1.460	14.76	23.86	30.24
Eva3	1225	23.93	1.470	18.21	23.86	30.21
Highmisp	2097	0.790	0.410	0	1	1
Commisp	2097	0.170	4.290	-18.95	1.070	3.180
Equity Financing	2124	0.400	0.490	0	0	1
Financing Efficiency	935	0.420	0.490	0	0	1
Property	2107	0.300	0.460	0	0	1
Profit	2097	10.84	10.96	-11.41	9.020	41.71
Capital Structure	2097	41.02	21.43	1.390	39.66	460.7
Operating Capacity	2097	0.750	0.530	0.0200	0.630	7.290
Corporate Governance	2092	0.370	0.0500	0	0.330	0.710
Acquisition Size	1924	0.210	0.350	0	0.0700	1.690
Rela	2016	0.440	0.500	0	0	1

#### 4.2 Empirical result analysis

##### 4.2.1 Equity financing, mispricing and M & A performance

Table 3 reports the moderating effect of the company's mispricing on the impact of equity financing on M & A performance, the regression results in Table 3 show that the coefficient of equity financing on the performance within three years after M&A are significantly positive, and equity financing helps increase M&A performance. The result is the same as the conclusions of Chen et al. (2020) that equity financing significantly increases the company's M&A performance[20].

However, after adding the cross multiplication term of overestimation and equity financing, it is found that the cross multiplication term coefficient is significantly negative, indicating that using equity financing when overestimated will reduce the positive impact of equity financing on M & A performance. It may be because when the company is overvalued, the company's managers will cater to the market for financing in order to maintain or increase the stock price, and raise more cheap capital, including equity financing(Jensen, 2005)[21], even if the company has sufficient internal funds(Polk & Sapienza, 2009)[22]. The internal governance structure of most listed companies in China is not perfect. For example, the state-controlled enterprises generally have the problems of the absence of major shareholders or high equity concentration, and the management lacks the initiative power of strategic M & A. Therefore, when the listed company is overestimated, it will conduct inefficient M&A financing out of catering to market motives, which can not help increase M & A performance.

**Table 3** The moderating effect of corporate mispricing on the impact of equity financing methods on M&A performance

Variables	(1)	(2)	(3)
	Eva1	Eva2	Eva3
Equity Financing	0.460*** (3.68)	0.460*** (3.23)	0.497*** (3.08)
Highmisp*Equity Financing	-0.317*** (-2.66)	-0.376*** (-2.77)	-0.440*** (-2.88)
Property	0.438*** (6.29)	0.461*** (5.69)	0.449*** (4.89)
Profit	0.034*** (10.91)	0.040*** (11.01)	0.038*** (9.08)
Capital Structure	0.019*** (12.09)	0.017*** (9.41)	0.015*** (7.65)
Operating Capacity	0.310*** (4.85)	0.330*** (4.46)	0.403*** (4.82)
Corporate Governance	-0.561 (-1.02)	-1.346** (-2.09)	-0.975 (-1.25)
Acquisition Size	0.696*** (7.07)	0.836*** (7.57)	0.615*** (5.09)
Rela	0.143** (2.16)	0.174** (2.21)	0.136 (1.46)
Constant	21.703*** (49.88)	22.557*** (47.10)	22.759*** (41.30)
Obs	1,702	1,378	1,051
R-squared	0.255	0.263	0.248
Ind	Yes	Yes	Yes
Year	Yes	Yes	Yes

Note: \*, \*\*, \*\*\*, denote significant at the levels of 10%, 5% and 1% respectively; T value is in parentheses; the same below.

#### 4.2.2 M&A company mispricing, financing efficiency and M&A performance

The above research found that using equity financing when overestimated reduces the positive impact of equity financing on M&A performance. The following explores the intermediary mechanism of timing financing on M & A performance from the perspective of financing efficiency. The regression results in Table 4 show that for path a, when M & A company is overestimated, the financing efficiency decreases. For path b, the reduction of merger financing efficiency will reduce the M & A performance, so the overestimation of M & A company will reduce M & A performance. For path c' , the overestimation of M&A companies reduces M & A performance. This conclusion supports the research conclusion of Polk and Sapienza (2009) that catering to the market in the long run when overestimated can not help increase value. We find path a and path b are both significant, so the mediating effect of financing efficiency is significant. Therefore, the intermediary effect of financing efficiency exists, that is, the overestimation of M & A companies affects M & A performance through financing efficiency.

**Table 4** Test of the mediating effect of M&A Financing Efficiency

Variables	(1)	(2)
	Fiance0	Eva1
Highmisp	-1.003*** (-3.971)	-0.927*** (-6.167)
Fiance0		0.269*** (3.240)
Property	-0.031 (-0.174)	0.294*** (3.302)
Profit	0.008 (0.724)	0.057*** (8.750)
Capital Structure	0.030*** (5.080)	0.023*** (8.218)
Operating Capacity	-0.301 (-1.599)	0.430*** (5.076)
Corporate Governance	0.543 (0.394)	0.409 (0.477)
Acquisition Size	0.406* (1.918)	0.702*** (6.078)
Rela	0.053 (0.321)	0.126 (1.458)
Constant	0.944 (0.753)	21.344*** (25.536)
Obs	859	857
R-squared		0.336
Ind	Yes	Yes
Year	Yes	Yes

## 5. Research Conclusions and Policy Implications

This study explains the impact mechanism of timing financing on M & A performance. It finds that increasing equity financing when overvalued reduces the positive impact of equity financing on M & A performance. Further research found that the mediating effect of the financing efficiency exists, that is, timing financing of listed companies reduces their financing efficiency, which can not help listed companies create value through M & A<sup>1</sup>.

Chinese companies face the dilemma of single financing channel at present. Therefore, it is necessary for listed companies to identify the mechanism of capital market mispricing affecting M & A performance, give full play to the favorable motivation of market value management, resist the unfavorable motivation of catering

<sup>1</sup> In order to ensure the robustness of the empirical results, we replaced the mispricing proxy variable with a continuous variable, it means defining  $\text{Commisp} = m_{it} - \bar{m}_{it}$  as the mispricing proxy variable. Then regress the model again, the research results are robust. Due to space limitations, it is no longer listed here.

to the market, and realize the synergistic growth of market value and intrinsic value through the combination of M & A and financing. Listed companies should also strengthen the concept of industry-finance integration and realize positive interaction between physical operation and capital operation of listed companies. M & A and financing using market timing should be based on company's overall value growth, and eliminate the high premium acquisition that can not create value.

## Reference

- [1] Li SM, Huang ZH, Guo JJ. Capital Market Pricing and Corporate Acquisitions: Evidence from Chinese A-share Firms. *Economic Research Journal*. 2020July; 55(7): 41-57.
- [2] Stein JC. Rational Capital Budgeting in an Irrational World. *Journal of Business*. 1996April; 69(4): 429-455.
- [3] Shleifer A, Vishny RW. Stock Market Driven Acquisitions. *Journal of Financial Economics*. 2003March; 70(3): 295-311.
- [4] Luo Q, Yang WY. Stock Mispricing and M&A : A Literature Review. *Journal of Beijing Technology and Business University(Social Sciences)*, 2019May; 34(05):81-91.
- [5] Shen D. Research on Financing Efficiency of SMEs in NEEQ Market: Private Placement Based on Three Stage DEA Model. *Audit & Economy Research*. 2017March; 32(3): 78-86.
- [6] Ying XJ, Liu XH, Li C. Analyze Financing Efficiency of Small and Micro Enterprises in Shandong Province—Base on DEA Model and Malmquist index. *Journal of Qingdao University (Natural Science Edition)*. 2018April; 31(4): 147-151.
- [7] Zhang HJ. Internal Control, Legal Environments and Corporate Financing Efficiency—Evidence from A-Share Listed Companies in China. *Journal of Shanxi Finance and Economics University*. 2017July; 39(7): 84-97.
- [8] Huang R, Ritter JR. Testing the Market Timing Theory of Capital Structure. *Kennesaw State: Kennesaw State University*. Working paper. 2005.
- [9] Alti A. How Persistent Is the Impact of Market Timing on Capital Structure?. *The Journal of Finance*. 2006April; 61(4): 1681-1710.
- [10] Hovakimian A. Are observed capital structures determined by equity market timing?. *Journal of Finance and Quantitative Analysis*. 2006January; 41(1): 221-243.
- [11] Kayhan A, Titman S. Firms' histories and their capital structures. *Journal of Financial Economics*. 2007January; 83 (1):1-32.
- [12] Vagenas NE. The Benefits of Overvaluation: Evidence from Mergers and Asquisitions. *Financial Management*. 2020January; 49(1),91-133.
- [13] Sun Y, Wei BY. Impact of Mispricing on the Consideration of Stock Payment: A Modified Model Based on Price-to-Book Ratio Decomposition. *Research on Financial and Economic Issues*. 2021March; (03):68-74.
- [14] Rhodes-Kropf M, Viswanathan S, Robinson DT. Valuation Waves and Merger Activity: The Empirical Evidence. *Journal of financial Economics*. 2005January; 77(1): 561-603.
- [15] Zuo XH, Lv JY. Research of Listed Companies' M&A Performance Based on EVA. *The Theory and Practice of Finance and Economics*. 2014June; 35(6): 34-39.
- [16] Zhao X, Li X, The Research of M&A Performance of Chinese Real Estate Industry Listed Companies. *Jilin Normal University Journal(Humanities & Social Science Edition)*. 2017May; 45(5): 94-103.
- [17] Shi GY, Liu GF. *Theory on Market Value Management*. Peking University Press.2008.
- [18] Wang XN, Du L, Wang QT. Analysis on financing efficiency of small and medium sized enterprises based on three stage DEA model. *Statistics and Decision*. 2016May; (5): 179-182.
- [19] Chen JQ, Zhao X, Niu XT. A Study on the Impact of Payment Methods M&A Performance from the Perspective of Financing Constrains. *Macroeconomics*. 2020April; (4):16-27.
- [20] Wen ZL, Ye BJ. Analyses of Mediating Effects: The Development of Methods and Models. *Advances in Psychological Science*. 2014May; 22(5): 731-745.
- [21] Jensen MC. Agency costs of overvalued equity. *Financial Management*. 2005January; 34(1): 5-19.
- [22] Polk C, Sapienza P. The Stock Market and Corporate Investment: A Test of Catering Theory. *Reriew of Financial Studies*. 2009January; 22(1): 187-217.

# Data Analytics Based Risk Management for Students' Performance – A Case Study

M. Somasundaram<sup>a,1</sup>, K.A. Mohamed Junaid<sup>a</sup>, D. Sudha<sup>b</sup>  
and Sabari L. Umamaheswari<sup>c</sup>

<sup>a</sup> Department of Electronics and Communication Engineering (ECE), R.M.K. Engineering College, Kavaraipettai, Tamilnadu 601206, India

<sup>b</sup> Department of Science & Humanities (S&H), R.M.K. Engineering College, Kavaraipettai, Tamilnadu 601206, India

<sup>c</sup> Department of Electrical and Electronics Engineering (EEE), R.M.K. Engineering College, Kavaraipettai, Tamilnadu 601206, India

**Abstract.** Business leaders around the world are using emerging technologies to capitalize on data, to create business value and to compete effectively in a digitally driven world. Among them the risk assessment and the risk management, based on the assessment is a process which can be made using the available past historical data and applying Data Analytics. Although it is being implemented in different business domains, it is at a nascent stage. It is further new and emerging in the area of Education. This paper describes such a process followed in an educational institution of an engineering college and the use of data for risk management. Based on the processes followed, the performance of the students is seen to be improving in academic performance, placement, higher education and entrepreneurship. This also provides a good process and framework for taking strategic initiatives which will give long term benefits in the areas like research and outreach activities.

**Keywords.** Data Analytics, Risk Management

## 1. Introduction

The risk assessment for an organization consists of using the past experience and the knowledge in identifying potential risks, and the possible impacts to the success of the business [1],[2] and [3]. This paper has implemented the Risk Analysis and Risk Management Plan at an institution level in an academic institution with focus on the students' performance in academic and career planning and settlement i.e. job placement, higher studies and entrepreneurship. Based on the processes followed, the performance of the students is seen to be improving in academic performance, placement, higher education and entrepreneurship. This also provides a good process and framework for taking strategic initiatives which will give long term benefits in the areas like research and outreach activities.

---

<sup>1</sup> Corresponding Author, Dr.M.Somasundaram, Professor, Department of Electronics and Communication Engineering, R.M.K. Engineering College, Kavaraipettai, Tamilnadu – 601206, India : E-mail: [mss.ece@rmkec.ac.in](mailto:mss.ece@rmkec.ac.in)

## 2. Methodologies for Data Analytics for Risk Management: Current status

The risk assessment for an organization consists of using the past experience and the knowledge in identifying potential risks, and the possible impacts to the success of the business [1],[2] and [3]. Big Data Analytics has been used for effective risk management in energy industry [4], banking [5], engineering and construction [6], business operations [7], and higher education [8]. However, in many such areas, as mentioned in their studies, it is at a nascent or early stage.

Typically, it consists of 6 steps: [9]

- Building a library of potential risks,
- Spotting test your data and validating your Key Risk Indicators (KRI),
- Connecting all of your data sources and automating testing,
- Digging deeper into the results and rounding out your analysis,
- Reporting and showcasing your work, and
- Expanding and repeating.

According to Mr. Vivek Katyal, principal, Deloitte & Touche LLP, Risk Information Services (RIS) practice leader and Audit and Enterprise Risk Services (AERS) leader for Deloitte Analytics, there is a resurgence of interest in the application of analytics to risk management challenges. [10].

## 3. Background of the Institutional Process

The institution, here presented as a case study, is an engineering college namely R.M.K. Engineering College (RMKEC) in India affiliated to Anna University, Chennai, India. RMKEC has been one of the top 10 institutions in Anna University in terms of academic excellence and has been achieving 95% consistently in the career settlement of students in terms of placement in jobs, higher studies and entrepreneurship. The college has 9 departments with a total number of students of over 3000. The details can be seen in the college website [11].

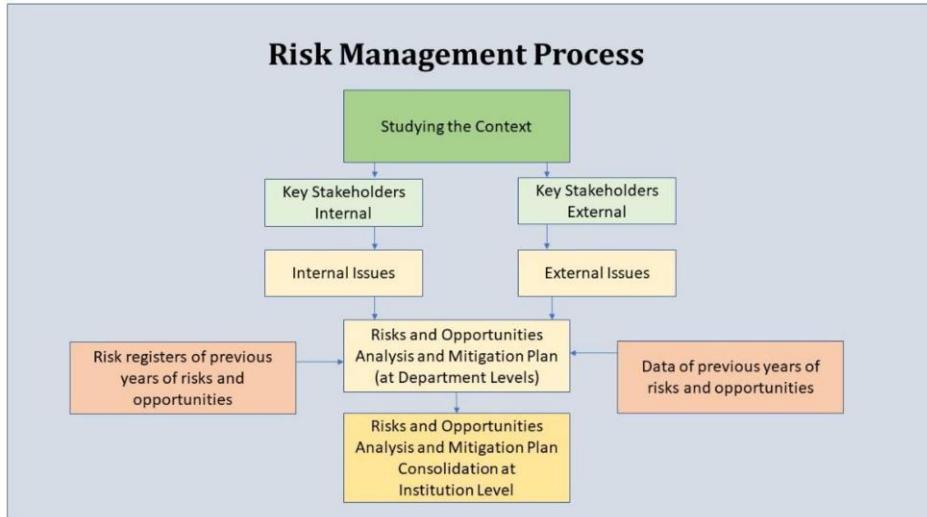
The Quality Management System (QMS) process presented has been implemented in the institution over 20 years since 1995. In view of the continuous improvement, the QMS Process has been upgraded in 2017 – 18 to consider a more formal way of addressing risks. The process was upgraded using the framework of ISO 9001: 2015 [12] which laid the processes for Risk Management. The Risk Management was done at the beginning of each academic year at department level as well as college level. Later all the risk management analysis are consolidated at college level and appropriate risk management activities are planned at college level and / or department level.

For the effective identification of risks, each department and the college analyses the data and thus the Data Analytics based Risk Management Process has been implemented in the last 2 years.

## 4. The Risk Management Process

The risk management process being followed as presented in the Figure 1. It has the steps as follows in line with the 6 steps given in [9]

The Risk Management process consists of steps as follows: Here this process is followed by each of the 8 departments of the college each having a Head of the Department and the faculty members of the department, each department having from 15 to 40 depending on the size of the department)



**Figure 1.** Risk Management Process

#### *Step 1: Building a library of potential risks*

When the Risk Management was implemented for the first time, a list was prepared by taking set of risks from existing literature and the past experiences in a qualitative approach. Subsequently as the process was implemented for 2 years, the set of risks identified have been included in the previous existing list.

#### *Step 2: Gathering and testing the background data and validating*

This is done by a process as follows:

**Studying the context** i.e., business and technical environment of the institution like ranking of the college in the industry, industry requirements of graduands, regulatory requirements like government regulations, University regulations, etc. Such data are collected over the period of time before the actual context study through stakeholders' inputs, updates from university, industry, professional bodies etc.

**Identification of key stakeholders' (both internal and external) expectations**, the current processes and their effectiveness of satisfying their expectations. In the case of the college, the internal stakeholders are students, faculty members and employees and external stakeholders are parents and alumni, industry (employers of students), regulatory bodies like University, All India Council for Technical Education (AICTE), Accreditation agencies like National Board of Accreditation (NBA), National Assessment and Accreditation Council (NAAC), etc.

Based on the expectations as above, the **internal and external issues** are identified. Some sample issues are given as illustrative here (not exhaustive).

**Table 1.** Sample Internal Issues:

Number	Internal Issues
1	Students' performance
2	Faculty competency
3	Infrastructure
4	Changing student culture

**Table 2.** Sample External Issues:

Number	External Issues
1	Technology changes needing knowledge update
2	Social media effecting student's concentration
3	Changing regulatory requirements
4	Economic changes affecting placement

*Step 3: Connecting all of them and analysing to identify the risks and opportunities*

Based on the above data collected and the issues identified, the **risks and opportunities** are identified by using the data as below:

**Table 3.** Data use for Analytics

Number	Area of Issues and Risk	Data Used for the Analytics
1	Academic performance	<ul style="list-style-type: none"> <li>• Evaluation Results and Causal Analysis for internal (College level) and external (University) assessments</li> <li>• Students University ranks of past students</li> </ul>
2	Placement	<ul style="list-style-type: none"> <li>• Job opportunities in the industry</li> <li>• Past placement data</li> <li>• Students Employability data (i.e. Aptitude, Soft skills etc) based on the tests conducted</li> </ul>
3	Innovation and Entrepreneurship	<ul style="list-style-type: none"> <li>• Number of startups</li> <li>• Number of innovative projects / competitions won</li> <li>• Number of events conducted on innovation and entrepreneurship</li> </ul>
4	Research	<ul style="list-style-type: none"> <li>• Number of papers published</li> <li>• Funds received for research</li> </ul>
5	Students Information	<ul style="list-style-type: none"> <li>• Students counseling data of personal information, mentoring details etc.</li> </ul>

**Table 4.** Risk Identification and their impact

Number	Risk	Event Likelihood (A)	Affecting Reputation and Business (B)	Risk Factor (A * B)
1	Technology changes needing knowledge update			
2	Social media effecting student's concentration			
3	Changing regulatory requirements			
4	Economic changes affecting placement			

**Note:**

**Event Likelihood (A)**

1. Rarely
2. Occasional
3. Likely
4. Frequent

**Affecting reputation and business (B)**

1. Minor
2. Moderate
3. Major
4. Critical

#### *Step 4: Further analysis of the risks and opportunities*

Further analysis is done and the risk mitigation plans are incorporated. A typical sample risk analysis consists of

- Risks and opportunities
- Steps taken by institution to address the risk / opportunity
- Rating of Risk Mitigation planned (Rating in a scale 1 to 10 : 10 effective; 1 not effective)

#### *Step 5: The Risk Consolidation*

The Risk Analysis and Mitigation plans are conducted by each of the 8 Departments. Then the Head of the Institution and Quality Management Team review them and consolidate into a single analysis as follows:

- R: Risk Factor : A \* B where A : Event Likelihood (1 low to 4 high ) and B : Affecting reputation and business (1 low to 4 high)
- M: Rating of risk mitigation in a scale of 1 to 10 (10 effective; 1 not effective)
- Note: Top 3 or 4 risks for each department are highlighted : Requires close monitoring and addressing the risks
- Note: Bottom 3 or 4 risk mitigation effectiveness ratings for each department are highlighted : Requires close monitoring and improving the effectiveness

A sample risk consolidation is prepared as a final summary for management review and planning

**Table 5.** Risk Consolidation Plan:

No	Risks / Opportunities	Dept 1		Dept 2		Dept 3		Dept 4	
		R	M	R	M	R	M	R	M
	<b>Risks (Examples, not exhaustive list)</b>								
1	Poor academic performance	1	8	12	9	12	9	2	8
2	Reduction in placement	4	7	8	8	8	8	4	7
3	Environment Factors (eg Rain, Floods, Pandemic Period )	2	9	2	6	2	6	9	6
20	Aspirants attracted to software courses							9	6
	<b>Opportunities (Examples, not exhaustive list)</b>								
1	Corporate contests eg SIH				9		9		8
2	Intra / Inter dept contests				9		9		8
3	Online certification				9		9		8
20	Online courses and practice for coding								

#### *Step 6: Communication and Action Planning*

Based on the consolidation, the Head of the Institution communicates the analysis to all the members of the departments and provides the action plans at the institution and department level. The action plans are implemented and monitored regularly on the status and effectiveness of the Risk Management plan.

## 5. Conclusion: Performance of the Students

The academic performance of the students have been consistent over the last 2 years of 85% pass percentage. However, the Risk Management has been effective in planning for risks due to COVID 19 in the year 2020 – 21. The institution switched over to online mode of teaching and learning and evaluation and effectively conducted the academic activities. It is to be noted that the institution was conferred the status of Autonomous by the Anna University with effect from the academic year 2020 – 21. The academic activities related to design of curriculum and syllabus were done effectively. The Autonomous status and the flexibility in designing the curriculum and syllabus will address some of the key risks of changes in technology and changes in the expectations of the industry.

The performance of the student in other key parameters are given as below. Further strategic initiatives have been taken to improve the students' performance on a long term of 2 to 3 years like research etc.

**Table 6.** Performance of the students: (Some key parameters)

N o	Key Performance Index	Year 2019 - 20	Year 2020 - 21	Strategic Initiatives Taken in 2020 - 21
1	No. of students placed in jobs	422	443	Note: Placement is still in progress
2	No. of students in High Salary packages	28	58	Identifying and training high performers
3	Startup companies by students	12	109	Established Institution's Innovation Council in 2019
4	Patents published / granted by students under guidance by faculty members	0	8	IPR Cell created

## 6. Future Direction

As we have initiated Data Analytics based Risk Management, we are currently implementing various tools in Artificial Intelligence and Data Science to consolidate the data and do analytics in terms of diagnosis, and prediction also.

## References

- [1] <https://www.henryharvin.com/blog/role-of-data-science-in-risk-management/> : Role of Data Science in Risk Management in 2021 [Updated]
- [2] <https://www.abs-group.com/Knowledge-Center/Insights/What-is-Data-Science-Applying-Analytics-to-Risk-Management/> What is Data Science? Applying Analytics to Risk Management
- [3] <https://financetrain.com/role-of-data-science-in-risk-management/> Role of Data Science in Risk Management
- [4] Application of big data analytics in process safety and risk management: Pankaj Goel, et al, 2017 IEEE International Conference on Big Data (Big Data), <https://ieeexplore.ieee.org/abstract/document/8258040>
- [5] Risk Management 4.0: The Role of Big Data Analytics in the Bank Sector: Dicuonzo, Grazia et al, International Journal of Economics and Financial Issues; Mersin Vol. 9, Iss. 6, (2019): 40-47. <https://www.proquest.com/openview/890e7032744535a4ff6f5c2f5efb079a/1?pq-origsite=gscholar&cbl=816338>

- [6] Data Analytics for Engineering and Construction Project Risk Management (2020) : Ivan Damnjanovic et al, ISBN: 978-3-030-14251-3 (2020) , Springer Link, <https://link.springer.com/book/10.1007%2F978-3-030-14251-3>
- [7] Recent Development in Big Data Analytics for Business Operations and Risk Management : Tsan-Ming Choi et al, IEEE Transactions on Cybernetics ( Volume: 47, Issue: 1, Jan. 2017), <https://ieeexplore.ieee.org/abstract/document/7378465>
- [8] Opportunities and challenges for big data analytics in US higher education: A conceptual model for implementation: Mohsen Attaran, et al, Sage Journals, April 26, 2018<https://journals.sagepub.com/doi/abs/10.1177/0950422218770937>
- [9] <https://www.wegalvanize.com/risk/data-analytics-risk-assessments/> : 6 steps to start using data analytics for risk assessments
- [10] <https://www2.deloitte.com/content/dam/Deloitte/au/Documents/risk/deloitte-au-risk-risk-angles-applying-analytics-risk-management-250215.pdf> : Deloitte : Risk Angles Five questions about applying analytics to risk management
- [11] [www.rmkec.ac.in](http://www.rmkec.ac.in) : R.M.K. Engineering College
- [12] <https://www.iso.org/standard/62085.html> : ISO 9001:2015 : Quality management systems — Requirements

# Comprehending the Pricing Decision of Online Car-Hailing Services in China: Price Regulation Vs Entry Limitation

Jie Yang<sup>a,1</sup> Daozhi Zhao<sup>b</sup> Zeyu Wang<sup>c</sup> Wanying Jiang<sup>d</sup>

<sup>a</sup> School of Social Science, Tsinghua University, 30 Shuangqing Road, Haidian District, Beijing, 100084, China

<sup>b</sup> College of Management and Economics, Tianjin University, 92 Weijin Road, Nankai District, Tianjin, 300072, China

<sup>c</sup> Department of Computer Engineering, Texas A&M University, 400 Bizzell St, College Station, TX, 77840, USA

<sup>d</sup> School Of Psychology, Northwest Normal University, 967 Anningxi Road, Anning District, Lanzhou, 730070, China

**Abstract.** Online car-hailing service has had an exponential growth in recent years, and poses a substantial threat to taxi service. Yet how to regulate online car-hailing service has not been adequately studied. Based on the regulation of tradition taxi service, price regulation and entry limitation are used to regulate online car-hailing service in this paper. Moreover, we consider two types of online car-hailing service, i.e., high-end service(e.g. UberBlack) and low-end service (e.g., UberX), according to the perceived service level. Then, the optimal platform's price is formulated. The result shows that the price regulation are likely to increase the optimal platform price, depending on the service type and the taxi price. When the platform offers low-end service and the taxi price is low, the optimal platform price does not change. In contrast, the entry limitation reduces the optimal platform's price when it achieves the regulation target.

**Keywords.** Sharing economy, online car-hailing service,taxi service, price regulation, entry limitation

## 1. Introduction

Facilitated by the information technology, many business operations are developing in online mode [1]. An example of such business models is the online car-hailing service, which enables riders to request rides online from independent drivers nearby [2]. It collects drivers offering ride services to the wider public based on coordination through a platform for centralized dispatch[3]. The popularity of online car-hailing service relies not only on advanced technology, but also on a lower price[4], higher flexibility[5], and environment friendly [6-7]. Therefore, the emergence of online car-hailing service presents a tremendous disadvantage to traditional taxi service. In China,

---

<sup>1</sup> Corresponding Author, Jie Yang, School of Social Science, Tsinghua University, 30 Shuangqing Road, Haidian District, Beijing,100084, China. E-mail: nemoyj1989@126.com.

the transaction volume of online car-hailing service in 2017 exceeded 200 billion with a potential increase [8].

The rapid development of online car-hailing platforms leads to the decline of taxi demand [9]. Correspondingly, the profit of taxi service is reduced. Taxi drivers are unsatisfied with the entry of online car-hailing service and claim that the government imposed different rules on the entrant than they must follow [10]. Thus, online car-hailing service has brought out emerging issues such as regulation. However, the research on the regulation policy is very few. This paper, motivated by the growth and the new rules of online car-hailing service, is focused on the two regulation policies, i.e., price regulation and entry limitation, according to game theoretic method of modeling. And we also characterize the two types of online car-hailing services by different perceived service levels compared to that of taxi service, i.e., high-end service and low-end service. On one hand, we aim to theoretically discuss about the impact of two regulation policies on the price strategy of the platform. On the other hand, we also hope to provide some suggestions about regulation of online car-hailing service for the government.

Our main findings are as follows. Our analysis indicates that the price regulation increases the optimal platform's price when the platform offers high-end service, whereas when the platform offers low-end service, the optimal platform's price does not change with and without regulation under certain circumstance. It is related to the taxi service price and the perceived service level. The entry limitation reduces the optimal price of online car-hailing service. The result shows that entry limitation benefits not only taxi service but also riders.

The remainder of this paper is organized as follows. In Section 2, we provide a review of related literature. In Section 3, we develop the model and analyze the model. In Section 4, we conduct the comparison analysis. We conclude the paper in Section 5.

## **2. Literature review**

This paper draws on three different research streams: studies about the taxi service, researches focused on the ridesharing service and the researches on the regulation of the transportation services.

[11] develops the first taxi model in an aggregated way, exploring the problem of the optimal price and the optimal service quality. Based on the model proposed by [11], [12] builds a model of taxi industry and determines equilibrium output and capacity in different market configurations. Then, researchers have been made to better understand and characterize the mechanism of how taxi demand and supply are equilibrated, and the impact of different market configurations; moreover, researchers have been incorporated the spatial structure of taxi services into models [13-16]. Since online car-hailing service resembles taxi service[9], we follow this stream to adopt the similar assumptions.

There is growing research that focuses on online car-hailing service, especially the platform pricing decision. Several papers study the platform price in static price policy or dynamic price policy [17-19]. Related research consider the riders and drivers behavior, such as delay sensitivity[20-21] and drivers idle-time sensitivity [22]. [17] suggests that dynamic policy can ensure a high service level in high demand periods. [23] argues that surge pricing can be a useful method in places where demand is even

lower than supply. For an extensive review on the research online car-hailing service, we refer readers to[24-25].

More relevant to this paper are literatures investigating the regulation of transportation service, especially the regulation of the ridesharing service. The regulation of taxi service focuses on price, quantity and quality [26-27]. [28] studies the impact of Uber on taxi service in the US and claims that cities with the taxicab medallion system should reduce the cost of taxi service when Uber enters the market. [29] considers the competition between two ridesharing service platforms and explores the effect of different regulation modes, such as price adjustments regulation and the joint with official platforms. Similarly, considering different regulation policies, [30] analyzes a two-period dynamic game to evaluate several performance measures such as consumer welfare and social welfare. Distinct from [29] and [30], we focus on the competition between online car-hailing service and taxi service, and we also consider two types of online car-hailing service according to the service experience. In addition, we attempted to compare the platform's price decision with different regulation policies, i.e., price regulation and entry limitation.

### 3. Model

#### 3.1. Monopoly model

The ride market considered in this paper includes taxi service and online car-hailing service. Other ride services, such as tubes and buses, are considered as outside options. Taxis play a significant paratransit role to other public transportation modes because of its flexibility and personalized service capabilities [31]. Taxi service was free of competition before the entry of online car-hailing service. Thus, we first analyze a basic riders' choice model in a market with taxi service only. Riders are heterogeneous and differ in their perceived value of ride services. Let  $v$ , which is a random number, denote how much a rider values taxi service. We assume that  $v$  follows uniform distribution,  $U(0, V)$ , and the total market size is normalized to one.

Ride service market is a dynamic time-varying system. We split time horizon (e.g., a whole day) into a set of periods to deal with such variances. When the duration of each period is sufficiently short, the market can be regarded as uniform in each period [32-33]. Note that we focus on the rider choice in one period excluding peak period and congestion period. We do not consider the case that demand exceeds supply. Moreover, Similar to [11], the trip length and the speed en route are assumed to remain constant in a single period within a geographic area. Therefore, the total cost of one trip can be considered as the service price [34]. The price of taxi service is strictly regulated by the government [35], who considers the public transport capacity and the volume of the road resource. Let  $\tilde{p}_T$  represent the given price of taxi service. We assume that riders are rational, that is, they make the decision to maximize their utilities. The utility of using taxi service is given by  $U_T = v - \tilde{p}_T$ . Suppose the utility is set to be 0 when the riders take outside options without loss of generality. Clearly, riders who gain positive utility will choose taxi service. Therefore, the demand of taxi service is

$$\lambda_T = \Pr(U_T \geq 0) = 1 - \frac{\tilde{p}_T}{V}.$$

### 3.2. Competition model without regulation

The online car-hailing platform, such as Uber and Didi, characterized by its innovative mobile internet technology and various service types, has experienced exponential growth in recent years. The platform matches riders with independent drivers in real time by smartphone applications [2]. Different from taxi service, the price of online car-hailing service is determined by the platform [35]. The platforms usually adopt low price strategy to attract riders [4]. Moreover, the platforms provide different services, i.e., high-end service (i.e., UberBLACK) and low-end service (i.e., UberX), according to the perceived service experience to better satisfy the riders' heterogeneous needs. In the competitive market, we assume that online car-hailing service is vertically differentiated with respect to taxi service. Let  $\alpha$  reflect the riders' differential preference. The rider value online car-hailing service as  $\alpha v$ . Specifically, we assume that the perceived service level of high-end service is higher than that of taxi service, i.e.,  $\alpha > 1$ . And the perceived service level of low-end service is lower than that of taxi service, i.e.,  $0 < \alpha < 1$ .

Let  $p_o$  represent the price of online car-hailing service. The riders' utility of choosing online car-hailing service is  $U_o(\alpha v, p_o) = \alpha v - p_o$ . Similar to Section 3.1, either taxi service or online car-hailing service that offers a rider greater positive utility will be chosen by the rider. Given the prices of taxi service and online car-hailing service, the demand of both services, i.e.,  $\lambda_T$  and  $\lambda_O$ , is shown in Table 1. Note that  $p_{OH}$  and  $p_{OL}$  denote the price of high-end service and low-end service, respectively.

Table 1 Demand of taxi service and online car-hailing service under different scenarios

Scenario	Condition	$\lambda_T$	$\lambda_O$
$\alpha > 1$	$0 < p_{OH} \leq \alpha \tilde{p}_T$	0	$1 - \frac{p_{OH}}{\alpha V}$
	$\alpha \tilde{p}_T < p_{OH} \leq \tilde{p}_T + \bar{\alpha}V$	$\frac{p_{OH} - \alpha \tilde{p}_T}{\bar{\alpha}V}$	$1 - \frac{p_{OH} - \tilde{p}_T}{\bar{\alpha}V}$
	$p_{OH} > \tilde{p}_T + \bar{\alpha}V$	$1 - \frac{\tilde{p}_T}{V}$	0
$0 < \alpha < 1$	$0 < p_{OL} < \alpha \tilde{p}_T + \bar{\alpha}V$	0	$1 - \frac{p_{OL}}{\alpha V}$
	$\alpha \tilde{p}_T + \bar{\alpha}V \leq p_{OL} < \alpha \tilde{p}_T$	$1 + \frac{\tilde{p}_T - p_{OL}}{\bar{\alpha}V}$	$\frac{p_{OL} - \tilde{p}_T}{\bar{\alpha}V} - \frac{p_{OL}}{\alpha V}$
	$p_{OL} \geq \alpha \tilde{p}_T$	$1 - \frac{\tilde{p}_T}{V}$	0

The platform's profit is given by  $\pi_o(p_o) = \lambda_o(p_o - c_o)$ , where  $c_o$  represents the operation cost. Given the price of taxi service, the optimal prices of high-end service and low-end service without regulation are derived as  $p_{NH}^* = \frac{\tilde{p}_T + c_{OH} + \bar{\alpha}V}{2}$  and

$p_{NL}^* = \frac{\alpha \tilde{p}_T + c_{OL}}{2}$ , where  $c_{OH}$  ( $c_{OL}$ ) is the operation costs of high-end (low-end) service.

### 3.3. Competition model with regulation

The rapid development of online car-hailing platforms leads to the decline of taxi demand due to the low price and shorter wait times [9,36]. Correspondingly, the profit of taxi service is reduced. According to [37], taxi drivers have reference-dependent preferences around an earned income target. Since the taxi price is fixed in this paper, taxi drivers have reference-dependent preferences around a demand target. A reference point divides outcomes into gains or losses, thus creating a qualitative difference in the valuation of outcomes slightly above or below that reference point [38]. [39] proposes that it is natural to assume that the reference group is the set of agents playing against each other. Therefore, we assume taxi drivers choose the demand of online car-hailing service as the reference point of demand. If the regulation can ensure the taxi demand satisfies the reference point of demand, i.e.,  $\lambda_T \geq \lambda_O$ , taxi drivers will accept the entry of online car-hailing service. Since online car-hailing service resembles taxi service [9] and regulation restrictions of taxi service include price regulation and entry limitation, we consider these two ways to ensure the coexistence of two services.

- Price regulation

Different from the case without regulation, the platform should consider the demand baseline of taxi drivers. Therefore, with price regulation, the price decision model of the platform is

$$\begin{aligned} \text{Max } & \pi_O(p_O) = \lambda_O(p_O - c_O) \\ \text{s.t. } & \lambda_O \leq \lambda_T \end{aligned}$$

**Proposition 2. (a).** The price of high-end service with price regulation should be in the interval  $[\frac{\bar{\alpha}V + (2 + \bar{\alpha})\tilde{p}_T}{2}, \tilde{p}_T + \bar{\alpha}V]$ . Moreover, the optimal price of high-end service with price regulation is  $p_{PH}^* = \frac{\bar{\alpha}V + (2 + \bar{\alpha})\tilde{p}_T}{2}$ , if  $c_T < \tilde{p}_T < V$ ; **(b).** The price of low-end service with price regulation should be in the interval  $[\frac{2\alpha\tilde{p}_T + \alpha\bar{\alpha}V}{(\alpha + 1)}, \alpha\tilde{p}_T]$ . Moreover, the optimal price of low-end service with price regulation,  $p_{PL}^*$ , is shown in Table 2.

Table 2 The optimal price of low-end service with price regulation

Condition of $\alpha$	Condition of $\tilde{p}_T$	$p_{PL}^*$
$\alpha \in (0, \tau]$	$\tilde{p}_T \in (\max\{c_T, \frac{c_{OL}}{\alpha}\}, \frac{(2\alpha\bar{\alpha}V - c_{OL}(\alpha + 1))}{\alpha(\alpha - 3)})$	$\frac{\alpha\tilde{p}_T + c_{OL}}{2}$
$\alpha \in (0, \tau]$	$\tilde{p}_T \in (\frac{(2\alpha\bar{\alpha}V - c_{OL}(\alpha + 1))}{\alpha(\alpha - 3)}, V)$	$\frac{2\alpha\tilde{p}_T + \alpha\bar{\alpha}V}{\alpha + 1}$
$\alpha \in (\tau, 1)$	$\tilde{p}_T \in (c_T, V)$	$\frac{2\alpha\tilde{p}_T + \alpha\bar{\alpha}V}{\alpha + 1}$

Note that  $\tau = \frac{c_{OL} + 2V - 3c_T + \sqrt{(c_{OL})^2 + (3c_T - 2V)^2 + 2c_{OL}(6V - 5c_T)}}{2(2V - c_T)}$ .

Proposition 2(a) indicates the price regulation always raises the price of high-end service, which is consistent with the result in the practice. The platform increases its price after the introduction of the new rules. Although online car-hailing service offers the service with a higher perceived service level, extravagant price compels riders who are indifferent between choosing taxi service or online car-hailing service in the absent of regulation to switch to taxi service. Proposition 2(b) indicates that under some circumstances, the optimal price of low-end service does not change with price regulation. Otherwise, the price regulation will raise the optimal price of low-end service. The condition is related to the taxi price and perceived service level coefficient,  $\alpha$ .

- Entry limitation

This paper focuses on the rider choice in one period excluding peak period and congestion period. Thus, the number of service providers can meet the demand as [11] does. The perceived service experience is based on the quality of drivers and convenience. However, considering the entry limitation, the number of service providers of online car-hailing service is reduced due to the restriction of the drivers' household registration and vehicle model. As a result, it increases the wait times of online car-hailing service. According to the data released by Didi, after the new rules launches, the difficulty of hailing a car at the high peak in Beijing, Shanghai, Guangzhou and Shenzhen in 2017 increases by 12.4%, 17.7%, 13.2% and 22.5%, comparing with the data of the same period in the previous year [40]. Thus, we assume the perceived service experience of online car-hailing service is reduced due to the curtailment of the service supply by entry limitation. Similar to [41], we also use  $\beta$  ( $0 < \beta < 1$ ) to denote the matching ratio of demand. It can also be referred as the regulation intensity. Then the utility of riders under entry limitation is modeled as  $U_O^E = \alpha\beta v - p_O$ . The price decision model of the platform is the same with Equation 1.

When the platform provides high-end service and the entry limitation is relatively loose ( $\frac{1}{\alpha} < \beta < 1$ ), such as the new rules in Jinan which has no definite restrictions on the driver's household registration and vehicle models, the optimal platform's price is similar to that in the price regulation. The feasible interval of the price of high-end service with entry limitation is  $p_{EH} \in [\frac{(\alpha\beta-1)V + (\alpha\beta+1)\tilde{p}_T}{2}, (\alpha\beta-1)V + \bar{p}_T]$ . The optimal price of high-end service with entry limitation is  $p_{EH}^* = \frac{(\alpha\beta-1)V + (\alpha\beta+1)\tilde{p}_T}{2}$ , if  $c_T < \tilde{p}_T < V$ .

If the entry limitation is so stringent that the perceived service experience of high-end services is reduced dramatically to be less than 1, i.e.,  $0 < \beta \leq \frac{1}{\alpha}$ . The optimal price of high-end service is shown in Table 3.

When the platform offers low-end service, the perceived experience of online car-hailing service is reduced from  $\alpha$  to  $\alpha\beta$  ( $0 < \beta < 1$ ) due to the entry limitation proposed by the government. The optimal price of low-end service with entry limitation is similar to that in the price regulation, substituting  $\alpha$  with  $\alpha\beta$ .

Table 3 The optimal price of high-end service with entry limitation when  $0 < \beta \leq \frac{1}{\alpha}$

Condition of $\beta$	Condition of $\tilde{p}_T$	$p_{EH}^*$
$\beta \in (0, \frac{\tau}{\alpha}]$	$\tilde{p}_T \in (\max\{c_T, \frac{c_{OH}}{\alpha\beta}\}, \frac{(2\alpha\beta(\alpha\beta-1)V - c_{OH}(\alpha\beta+1))}{\alpha\beta(\alpha\beta-3)}]$	$\frac{\alpha\beta\tilde{p}_T + c_{OH}}{2}$
$\beta \in (0, \frac{\tau}{\alpha}]$	$\tilde{p}_T \in (\frac{(2\alpha\beta(\alpha\beta-1)V - c_{OH}(\alpha\beta+1))}{\alpha\beta(\alpha\beta-3)}, V)$	$\frac{2\alpha\beta\tilde{p}_T + \alpha\beta(\alpha\beta-1)V}{\alpha\beta+1}$
$\beta \in (\frac{\tau}{\alpha}, 1)$	$c_T < \tilde{p}_T < V$	$\frac{2\alpha\beta\tilde{p}_T + \alpha\beta(\alpha\beta-1)V}{\alpha\beta+1}$

#### 4. Comparison analysis

The comparison of the optimal platform's price and the demand of taxi service between the price regulation and the entry limitation when the platform provides high-end service is shown in Table 4.

Table 4 The comparison result of the optimal price and the demand between two regulation policies when  $\alpha > 1$

Regulation intensity	$\tilde{p}_T$	The comparison of optimal price	The comparison of taxi service's demand
$\frac{1}{\alpha} < \beta < 1$	$\tilde{p}_T \in (c_T, \frac{\bar{\alpha}V + c_{OH}}{2\alpha-1})$	$p_{EH}^* < p_{NH}^* < p_{PH}^*$	$\lambda_T^E = \lambda_T^P > \lambda_T^N$
$\frac{1}{\alpha} < \beta < 1$	$\tilde{p}_T \in (\frac{\bar{\alpha}V + c_{OH}}{2\alpha-1}, V)$	$p_{EH}^* < p_{PH}^*$	$\lambda_T^E = \lambda_T^P$
$\frac{\tau}{\alpha} < \beta < \frac{1}{\alpha}$	$\tilde{p}_T \in (c_T, V)$	$p_{EH}^* < p_{NH}^* < p_{PH}^*$	$\lambda_T^E > \lambda_T^P > \lambda_T^N$
$0 < \beta < \frac{\tau}{\alpha}$	$\tilde{p}_T \in (\max\{c_T, \frac{c_{OH}}{\alpha\beta}\}, V)$	$p_{EH}^* < p_{NH}^* < p_{PH}^*$	$\lambda_T^E > \lambda_T^P > \lambda_T^N$

Note that  $\lambda_T^N$ ,  $\lambda_T^P$  and  $\lambda_T^E$  represent the taxi demand in the market without the regulation, with price regulation and with entry limitation.

From Table 4, the optimal price with price regulation is the highest, and the optimal price with entry limitation is the lowest. This result contradicts the fact that the new rules focused on entry limitation increase the price of online car-hailing service. In this paper, we consider the impact of entry limitation on the riders' utility, which will reduce the demand. Whereas the demand in the practice usually is unchanged or change only slightly when the service supply decreases. Note that the reduction of drivers influences the perceived service experience of riders. For example, riders may wait for a longer time for the car's arrival. Therefore, the platform attracts riders by reducing the price. The entry limitation achieves the coexistence of the two services and enables riders to enjoy a higher perceived service experience with a lower price simultaneously.

Comparing the optimal platform's prices with price regulation and entry limitation when the platform offers low-end service, we obtain the Proposition 3.

**Proposition 3** When the regulation intensity is sufficiently stringent ( $0 < \beta < \frac{\tau}{\alpha}$ ), there exists  $p^1 \in (\max\{c_T, \frac{c_{OL}}{\alpha\beta}\}, V)$  such that  $p_{NL}^* = p_{PL}^* > p_{EL}^*$ ,  $\lambda_T^E > \lambda_T^P = \lambda_T^N$  if  $\tilde{p}_T \leq p^1$ , and  $p_{PL}^* > p_{NL}^* > p_{EL}^*$ ,  $\lambda_T^E > \lambda_T^P > \lambda_T^N$  if  $\tilde{p}_T > p^1$ ; when the regulation intensity is sufficiently loose ( $\frac{\tau}{\alpha} < \beta < 1$ ),  $p_{PL}^* > p_{NL}^* > p_{EL}^*$ ,  $\lambda_T^E > \lambda_T^P > \lambda_T^N$ , where  $p^1 = \frac{(2\alpha\bar{\alpha}V - c_{OL}(\alpha+1))}{\alpha(\alpha-3)}$ .

Proposition 3 shows that the entry limitation leads to the lowest price of low-end service and the highest demand of taxi service regardless of the regulation intensity and the taxi service price. Moreover, the price regulation can result in the price of low-end service and the demand of taxi service no less than that in the deregulation market. Different from the price regulation, the platform chooses a lower price strategy to deal with the entry limitation. Meanwhile, the reduction of price depends on the regulation intensity. To be specific, when the regulation intensity is sufficiently stringent and the price of taxi service is lower, the optimal platform's price with price regulation equals to that without regulation. However, the entry limitation leads to a lower price, which is beneficial to riders. Note that the entry limitation incurs a sharp slump in the demand of online car-hailing service, although the price is lower.

## 5. Conclusion

In this paper we explore the pricing decision of online car-hailing service with different regulation policies, i.e., price regulation and entry limitation. We compared the optimal platform's price and the demand of taxi service under these three scenarios: without regulation, with price regulation and with entry limitation. Our findings indicate that the price regulation increases the platform's price when the platform provides high-end service. Whereas the optimal platform's price of low-end service is the same with and without price regulation under certain circumstance. It is related to the taxi service price and the perceived service experience. The entry limitation reduces the optimal platform's price to achieve the regulation target. The result implies that entry limitation benefits not only riders but also taxi drivers. In this paper, we considered the platform as the decision maker; however, the online car-hailing service has the features of a two-sided market. Drivers, self-employment workers, are entitled to decide when to work. Therefore, exploring the impact of new rules on performance of online car-hailing service based on two-sided market theory may be a possible direction for future research.

## Acknowledgement

This work is supported by National Natural Science Foundation of China (NSFC) grant[grant number 72072125].

## References

- [1] Raju V. Implementing flexible systems in doctoral viva defense through virtual mechanism. *Global Journal of Flexible Systems Management*. 2021 Jun;22(2):127-39.
- [2] Conway MW, Salon D, King DA. Trends in taxi use and the advent of ridehailing, 1995–2017: Evidence from the US National Household Travel Survey. *Urban Science*. 2018 Sep;2(3):79.
- [3] Lavieri PS, Dias FF, Juri NR, Kuhr J, Bhat CR. A model of ridesourcing demand generation and distribution. *Transportation Research Record*. 2018 Dec;2672(46):31-40.
- [4] Kortum K. New TRB special report: Between public and private mobility: Examining the rise of technology-enabled transportation services. *TR News*. 2016 Mar(302).
- [5] Pham A, Dacosta I, Jacot-Guillarmod B, Huguenin K, Hajar T, Tramèr F, Gligor V, Hubaux JP. Privateride: A privacy-enhanced ride-hailing service. *Proceedings on Privacy Enhancing Technologies*. 2017 Apr;2017(2):38-56.
- [6] Marić J, Opazo-Basáez M. Green servitization for flexible and sustainable supply chain operations: A review of reverse logistics services in manufacturing. *Global Journal of Flexible Systems Management*. 2019 Dec;20(1):65-80.
- [7] Zhong J, Lin Y, Yang S. The Impact of Ride-Hailing Services on Private Car Use in Urban Areas: An Examination in Chinese Cities. *Journal of Advanced Transportation*. 2020 Nov 23;2020.
- [8] Deloitte. Market Analysis of Car-hailing Business. <https://www2.deloitte.com/content/dam/Deloitte/cn/Documents/strategy/deloitte-cn-consulting-online-car-hailing-en-190221.pdf>.2017.
- [9] Jiang W, Zhang L. The impact of the transportation network companies on the taxi industry: Evidence from Beijing's GPS taxi trajectory data. *Ieee Access*. 2018 Feb 27;6:12438-50.
- [10] Cetin T, Deakin E. Regulation of taxis and the rise of ridesharing. *Transport Policy*. 2019 Apr 1;76:149-58.
- [11] Douglas GW. Price regulation and optimal service standards: The taxicab industry. *Journal of Transport Economics and Policy*. 1972 May 1:116-27.
- [12] De Vany AS. Capacity utilization under alternative regulatory restraints: an analysis of taxi markets. *Journal of political economy*. 1975 Feb 1;83(1):83-94.
- [13] Sayarshad HR, Gao HO. A scalable non-myopic dynamic dial-a-ride and pricing problem for competitive on-demand mobility systems. *Transportation Research Part C: Emerging Technologies*. 2018 Jun 1;91:192-208.
- [14] Wang J, Bai Z, Hu X. The effect of the integrated service mode and travel time uncertainty on taxis network equilibrium. *Mathematical Problems in Engineering*. 2015 Jan 1;2015.
- [15] Yang H, Fung CS, Wong KI, Wong SC. Nonlinear pricing of taxi services. *Transportation Research Part A: Policy and Practice*. 2010 Jun 1;44(5):337-48.
- [16] Wong RC, Yang L, Szeto WY, Li YC, Wong SC. The effects of accessible taxi service and taxi fare subsidy scheme on the elderly's willingness-to-travel. *Transport Policy*. 2020 Oct 1;97:129-36.
- [17] Cachon GP, Daniels KM, Lobel R. The role of surge pricing on a service platform with self-scheduling capacity. *Manufacturing & Service Operations Management*. 2017 Jul;19(3):368-84.
- [18] Sun L, Teunter RH, Babai MZ, Hua G. Optimal pricing for ride-sourcing platforms. *European Journal of Operational Research*. 2019 Nov 1;278(3):783-95.
- [19] Ma J, Xu M, Meng Q, Cheng L. Ridesharing user equilibrium problem under OD-based surge pricing strategy. *Transportation Research Part B: Methodological*. 2020 Apr 1;134:1-24.
- [20] Bai J, So KC, Tang CS, Chen X, Wang H. Coordinating supply and demand on an on-demand service platform with impatient customers. *Manufacturing & Service Operations Management*. 2019 Jul;21(3):556-70.
- [21] Benjaafar S, Kong G, Li X, Courcoubetis C. Peer-to-peer product sharing: Implications for ownership, usage, and social welfare in the sharing economy. *Management Science*. 2019 Feb;65(2):477-93.
- [22] Taylor TA. On-demand service platforms. *Manufacturing & Service Operations Management*. 2018 Oct;20(4):704-20.
- [23] Guda H, Subramanian U. Your uber is arriving: Managing on-demand workers through surge pricing, forecast communication, and worker incentives. *Management Science*. 2019 May;65(5):1995-2014.
- [24] Furuhata M, Dessouky M, Ordóñez F, Brunet ME, Wang X, Koenig S. Ridesharing: The state-of-the-art and future directions. *Transportation Research Part B: Methodological*. 2013 Nov 1;57:28-46.
- [25] Benjaafar S, Hu M. Operations management in the age of the sharing economy: what is old and what is new?. *Manufacturing & Service Operations Management*. 2020 Jan;22(1):93-101.
- [26] Schroeter JR. A model of taxi service under fare structure and fleet size regulation. *The Bell Journal of Economics*. 1983 Apr 1:81-96.
- [27] Harding S, Kandlikar M, Gulati S. Taxi apps, regulation, and the market for taxi journeys. *Transportation Research Part A: Policy and Practice*. 2016 Jun 1;88:15-25.

- [28] Bond AT. An app for that: Local governments and the rise of the sharing economy. *Notre Dame L. Rev. Online.* 2014;90:77.
- [29] Wang S, Chen H, Wu D. Regulating platform competition in two-sided markets under the O2O era. *International Journal of Production Economics.* 2019 Sep 1;215:131-43.
- [30] Yu JJ, Tang CS, Max Shen ZJ, Chen XM. A balancing act of regulating on-demand ride services. *Management Science.* 2020 Jul;66(7):2975-92.
- [31] Szeto WY, Wong RC, Yang WH. Guiding vacant taxi drivers to demand locations by taxi-calling signals: A sequential binary logistic regression modeling approach and policy implications. *Transport Policy.* 2019 Apr 1;76:100-10.
- [32] Gan J, An B. Game theoretic considerations for optimizing efficiency of taxi systems. InWorkshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence 2015 Apr 1.
- [33] Gan J, An B, Wang H, Sun X, Shi Z. Optimal pricing for improving efficiency of taxi systems. InTwenty-Third International Joint Conference on Artificial Intelligence 2013 Jun 30.
- [34] Kim YJ, Hwang H. Incremental discount policy for taxi fare with price-sensitive demand. *International Journal of Production Economics.* 2008 Apr 1;112(2):895-902.
- [35] Taylor BD, Chin R, Crotty M, Dill J, Hoel LA, Manville M, Polzin S, Schaller B, Shaheen S, Sperling D. Between Public and Private Mobility: Examining the Rise of Technology-Enabled transportation Services; Special Report 319. Transportation Research Board: Committee for Review of Innovative Urban Mobility Services: Washington, DC, USA. 2015.
- [36] Jalloh, M. Uber: Advantages and Disadvantages. Investopedia. 2021.<https://www.investopedia.com/articles/investing/110614/taxi-industry-pros-cons-uber-and-other-ehail-apps.asp>
- [37] Camerer C, Babcock L, Loewenstein G, Thaler R. Labor supply of New York City cabdrivers: One day at a time. *The Quarterly Journal of Economics.* 1997 May 1;112(2):407-41.
- [38] Allen EJ, Dechow PM, Pope DG, Wu G. Reference-dependent preferences: Evidence from marathon runners. *Management Science.* 2017 Jun;63(6):1657-72.
- [39] Fehr E, Schmidt KM. A theory of fairness, competition, and cooperation. *The quarterly journal of economics.* 1999 Aug 1;114(3):817-68.
- [40] Wei, X., Han, Z, Peng, C. The new policy of online car-hailing service is implemented for one year: it is difficult to "make a comeback" by taxi? *Economic Information Daily.* 2017. [http://www.jjckb.cn/2017-07/28/c\\_136478886.htm](http://www.jjckb.cn/2017-07/28/c_136478886.htm)
- [41] Bellos I, Ferguson M, Toktay LB. The car sharing economy: Interaction of business model choice and product line design. *Manufacturing & Service Operations Management.* 2017 May;19(2):185-201.

# A Comparative Evidence of Income Levels Reflecting Gen Z's Digital Payments Intention and Usage

Khwanjira PONSREE<sup>a,b</sup>, Taksin PHONGPAEW<sup>b</sup>  
and Phaninee NARHUTARADHOL<sup>a,bl</sup>

<sup>a</sup> EXPEDITE / Global Management Consulting Center, Khon Kaen University,  
Thailand

<sup>b</sup> International College, Khon Kaen University, Thailand

**Abstract.** The study investigating e-payment services intention and use behavior of generation Z – knowing as the digital native cohort. To better understand the effect of financial volatility, the present study has categorized the study groups into low-income and high-income levels in accordance with Thai minimum wages. The UTAUT theory has been adopted to examining Gen Z's intention and behavior toward using e-payment services. The factors comprising of performance expectancy, effort expectancy, social influence expectancy, facilitating conditions affecting behavioral intention and use behavior. Thus, there was five main hypotheses have developed based on the utilized theory. The findings illustrate the significant differences between the two study groups. The behavior of using e-payment services is significantly distinct by the high-income level of Gen Z would rather using the services than the low-income. For Gen Z who earn a low income, there is one hypothesis that is rejected – effort expectancy affecting behavioral intention. On the other hand, for the higher-income, two hypotheses have not been confirmed which are performance expectancy and social influence expectancy positively influence behavioral intention. Among those supported hypotheses, the effect of intention on use behavior is the strongest path relationship for both low-income ( $\beta = 0.761$ ) and high-income levels ( $\beta = 0.576$ ).

**Keywords.** E-payment, generation Z, income levels, technology adoption, UTAUT

## 1. Introduction

Recently, the world is facing the unprecedented time of Covid-19 pandemic causing the economic shock-waves, exacting an enormous human toll, changing the way of living and livelihood, as well as shutting down the world communities, continents, regions, countries, cities, and societies [1]. After the outbreak of the Covid-19, Asian economies and Asian households were affected by various perspectives from the policy interventions of lockdowns, social distancing, as well as the restriction in tourism [2]. Among the observed Asian countries, including Cambodia, Indonesia, the Lao PDR, Malaysia, Myanmar, the Philippines, Thailand, and Viet Nam, about three quarters are

---

<sup>1</sup>Corresponding author, Phaninee NARHUTARADHOL, Global Entrepreneurship Development Center, Khon Kaen University, 123 Moo 16 Mitraphap Rd., Nai-Muang, Muang District, Khon Kaen, 4002, Thailand; E-mail: Phaninee@kku.ac.th

experiencing the reduction of households income, and almost all households experiencing financial illiquidity and weakness leading to reduce unnecessary consumptions [2]. United Nations Conference on Trade and Development (UNCTAD) argued that even though the Covid-19 crisis has destroyed the worlds' economies, the digital transformation has been accelerated at the same time [3]. Those policies and activities would create a long-term impact for the e-commerce market to have the potential to grow, especially in developing countries [3]. Capgemini Research Institute revealed that after the Covid-19 outbreak, people are preferable to use non-cash transactions instead of physical cash. The increase in non-cash transactions in 2019 rise 14% (worth USD 708.5 billion), the highest growth in the past decade. In 2019, the major growth is in the Asia Pacific, Europe, and North America grew by 25% (USD 243.6 billion) approximately resulting from the adoption of mobile payments and digital wallets widely in the region [4].

As the evidence of growth in e-commerce and e-payment during the Covid-19 pandemic which is contrary to the global economy, we see the intention and behavior of consumers or users are an interesting topic to study, therefore, the unified theory of acceptance and use of technology (UTAUT) has used to examine. Furthermore, generation Z is different from the other generational cohorts in the way of behavior, characteristics, preferences, and attitudes toward technology [5]. With the household financial volatility during an unprecedented time of Covid-19 pandemic [1], the accelerated time of digital transformation [3], and digitally native era [5], this present study see the opportunity to assess the moderating effect of income level in the household's financial volatility situation which caused by the Covid-19 crisis on e-payment intention of generation Z.

## 2. Literature Review

### 2.1. Generation Z

Generation Z or Gen Z have been defined as a digitally native generation who are most familiar with the internet and technology when compare with other generational cohorts [5]. Besides, Oxford Economics claimed that Gen Z is the generation who were born around the mid-1990s to 2010 and now are about one-third of the global population which have a high potential to drive the global economies in the near future [5]. Some researchers unveiled the Gen Z that they are independent and different from millennials because they seem to be interested in startup and early-stage companies as their career development paths [6]. Also, when comparing Gen Z with other cohorts, Gen Z at an individual level are most likely to accept the diversities regardless of races, gender, and orientations – the non-binary or third gender are accepted [6].

### 2.2. Income Level

The low-income segment has consisted of a group of people who spent daily life about USD 3 to USD 5 a day [7], on the other hand, the United Nations Development Program (UNDP) defined the low-income level as a group living on lower than USD 8 a day [8]. Nevertheless, Many scholars further studied the lower-income consumers – defined as the bottom of the pyramid (BOP) segment, whose living costs are beyond USD 10 a day [9]. In this present study, the income level has been categorized into low- and high-

income levels based on the average minimum wages rate of Thai labor which is THB 321 a day or THB 9,633 a month (THB = Thai Baht) [10]. However, THB 10,000 a month is more general. Therefore, this present study defines the low-income level would be under THB 10,000 a month while the high-income level would be upper THB 10,000 a month.

### *2.3. The Unified Theory of Acceptance and Use of Technology*

The current study has adopted the UTAUT model to examine Gen Z user's e-payment intention and behavior in Thailand. The constructs including; performance expectancy (PE) is the degree to which individual believes that the technology would enhance their job performance to be accomplished efficiently and PE has a positive effect on consumers' behavioral intention; effort expectancy (EE) has defined as the easiness of using technology that would not be complex and difficult and EE had a positively significant effect on behavioral intention; social influence expectancy (SIE) is defined as the individual perception of other people, surrounded people, or their important people expect them to use the new technology or system which has a significant impact on behavioral intention; facilitating conditions (FC) refer to consumers' belief of there are resources and support that exist for them to study and learn the way of using new technology or system and have a direct effect to use behavior; and behavioral intention (BI) of consumers creates a great effect and reflects the technological adoption, and their use behavior (UB) [11]. According to the aforesaid review, the path relationships and hypotheses have been demonstrated in table 1.

## **3. Methodology**

The study focusing on study Gen Z's intention and behavior of using e-payment in Thailand. The participants who age between 18 – 25 years old. The screen question asking about their experience of using e-payment services at least one time to ensure that we access our targeted participants. The duration of data collection is in the first half of the year 2021. During the Covid-19 pandemic, we use an online survey by applying Google Forms as the data collection tool.

The questions within the questionnaire are adapted from the previously published works of literature including three questions for each factor comprised of PE, EE, SIE, and FC [11,12], four questions for BI [11–13], and three questions for UB [11,13]. There are four parts to the questionnaire: (i) demographic information, (ii) background information of using e-payment services, (iii) nineteen questions regarding the factors potentially affect e-payment intention and usage where the respondents are required to rank the score from 1 to 5 based on Five-Point Linker-Scale [14], and (iv) the suggestions.

In this present study, Cochran's sample size determination has been used in case of the population of Gen Z is unknown [15] (there was insufficient data of Gen Z's population in Thailand). With the randomized population's proportion was estimated at 50%, 95% confidential interval, and 5% margin of error, therefore the approximate sample size is 384. The appropriate sample size when applying structural equation modeling (SEM) is about 250 to 500 which helps to eliminate the unexpected issues in the data analysis process [16]. Thus, the data from 500 respondents have been collected for the study and usable samples are 476.

## 4. Research Findings

The data have been collected from Gen Z who age between 18 – 25 years old and familiar with e-payment services. Females and males are accounted for 61.38% and 32.98% respectively. The main participants are the student (91.60%) following by the minority groups of employees (5.04%) and business owners (2.10%). The lower-income group accounted for 75% and the rest is the high-income group (25%). The t-test result shows a significant difference among the low- and high- income groups toward e-payment usage of Gen Z ( $t\text{-value} = -8.594$ , critical value = 1.971, d.f. = 217, and  $p\text{-value} < 0.001$ ). More than half of the high-income group has used e-payment services more than 20 times a month. In contrast, around two-third of the low-income group has used e-payment services lower than 10 times a month. Furthermore, we have assessed the normality of the collected data by measure mean, standard deviations, skewness, and kurtosis. The range of mean is between 4.070 (SIE3) to 4.470 (PE3). The standard deviation ranges from 0.727 (PE1) to 0.945 (SIE3). To test the univariate normality for each variable and the symmetry of deviation, skewness, and kurtosis are necessary to measure [17]. To represent normality, the lower skewness value of 2 and kurtosis value of 7 is required [18]. The skewness value is between -1.223 (PE construct) to -0.950 (BI construct) and the kurtosis value is in the range of 0.352 (BI construct) and 1.144 (PE construct). Thus, the data has represented the normality.

### 4.1. Confirmatory Factors Analysis

In the CFA process, the model fit of the measurements, convergent validity, and discriminant validity have been measured [19]. To confirm the validity and reliability of the overall measurement model, the model is required to achieve all thresholds including chi-square value or  $\chi^2$  ( $p < 0.05$ ),  $\text{CMIN}/\text{df} < 3.00$ ,  $\text{RMR} < 0.08$ ,  $\text{AGFI} > 0.80$ ,  $\text{PGFI} > 0.50$ ,  $\text{IFI} > 0.90$ ,  $\text{TLI} > 0.90$ ,  $\text{CFI} > 0.90$ , and  $\text{RSMEA} < 0.08$  [20]. Hence, the result achieved the threshold by chi-square value or  $\chi^2 = 360.17$  ( $p < 0.05$ ),  $\text{CMIN}/\text{df} = 2.629$ ,  $\text{RMR} = 0.021$ ,  $\text{AGFI} = 0.898$ ,  $\text{PGFI} = 0.668$ ,  $\text{IFI} = 0.971$ ,  $\text{TLI} = 0.964$ ,  $\text{CFI} = 0.971$ , and  $\text{RSMEA} = 0.059$ .

The convergent validity has performed by analyzing various measurements including; the test of factor loading to confirms the relation of variables within factor and the value of each item are required to be greater than 0.5 [20] – the values of factor loading within constructs have met the criteria of all items from 0.743 (FC3) to 0.914 (BI2); the average variance extract (AVE) is the examination of construct variance relative to the measurement error variance which the recommended value of AVE is over 0.5 to indicate the usability of the tested model [21] – The AVE value of all constructs are represent the usability including PE = 0.738, EE = 0.703, SIE = 0.776, FC = 0.674, BI = 0.783, and UB = 0.693; and composite reliability (CR) use for testing the internal consistency of the items within the construct, the higher value than 0.7 representing high consistent [20], hence, all constructs have a high internal consistency including PE = 0.894, EE = 0.876, SIE = 0.912, FC = 0.860, BI = 0.935, and UB = 0.871. The values of the factor loading, AVE, and CR achieved the aforesaid threshold, therefore the structural model has achieved convergent validity.

Discriminant validity is a crucial measurement to identify the distinction of one construct from another's constructs whether homogeneous or heterogeneous [22]. With the recommendation from prior studies, a chi-square different test has been used to validate [23,24]. They have suggested comparing the constrained and unconstrained

models of each pair of latent variables. All constructs were paired against one another's, thus there are 15 pairs in total. The correlation of two latent variables is freely correlated for the unconstrained model, on the other hand, fixing the correlation at one for the constrained model. Either chi-square different of higher than 3.84 or the different of degree of freedom of 1 (p-value < 0.05) indicating discriminant validity (p < 0.05) [23,24]. After the examination, the discriminant validity of the latent variables has been confirmed.

#### 4.2. Structural Equation Modeling

Structural equation modeling (SEM) is the second step after validating the model and CFA, which SEM uses for testing the hypothesis [19,25,26]. The present study utilizing Amos version 26 to analyze a crucial effect in which examining on multigroup analysis of income levels. The structural model is fit with the thresholds of all aspects [20] including chi-square value or  $\chi^2 = 675.874$  (p < 0.05), CMIN/df = 2.397, RMR = 0.029, AGFI = 0.829, PGFI = 0.648, IFI = 0.949, TLI = 0.938, CFI = 0.949, and RSMEA = 0.054.

**Table 1.** Hypothesis testing results of the multigroup analysis.

<b>Low-Income Level</b>					
Hypothesis	Effect	Standardized coefficient (b)	t-value	p-value	Results
H1a	PE → BI	0.445	2.673	0.008**	Supported
H2a	EE → BI	0.143	0.892	0.372	Not supported
H3a	SIE → BI	0.232	5.071	***	Supported
H4a	FC → UB	0.312	5.107	***	Supported
H5a	BI → UB	0.761	12.045	***	Supported
<b>High-Income Level</b>					
Hypothesis	Effect	Standardized coefficient (b)	t-value	p-value	Results
H1b	PE → BI	0.106	0.423	0.672	Not supported
H2b	EE → BI	0.732	2.564	0.010*	Supported
H3b	SIE → BI	-0.013	-0.147	0.883	Not supported
H4b	FC → UB	0.469	4.337	***	Supported
H5b	BI → UB	0.506	5.761	***	Supported

Structural wright: Degree of freedom = 18; CMIN = 60.957; and p-value < 0.001. Denotes that: \*\*\*p-value < 0.001, \*\*p-value < 0.01, and \*p-value < 0.05.

As illustrated in table 1, based on structural weight measurement, the value of CMIN is 60.957 with a degree of freedom of 18 (p-value < 0.001) indicates a significant difference among the two study groups. On the other hand, there are some similar and different results of path relationships. For Gen Z who earn a low income, there was the confirmation that PE reflected their BI (H1a;  $\beta = 0.445$ , t-value = 2.673, p < 0.01), contrary, the result of high-income Gen Z is rejected (H1b;  $\beta = 0.106$ , t-value = 0.423, p = 0.672). EE has not influenced BI of Gen Z who earn low income (H2a;  $\beta = 0.143$ , t-value = 0.892, p = 0.372) but it influences the BI of high-income (H2b;  $\beta = 0.732$ , t-value = 2.564, p < 0.05). To discuss at this point, since Gen Z are the digitally native cohort, they have most familiar with electronic devices and online activities by using the internet [5]. The ease of using e-payment had no significant effect on their intention, especially for the low-income group. In contrast, the higher-income group which most of them are heavy users would tend to prefer the easy method of using e-payment services because they have used the service repeatedly.

In other perspectives, there was a confirmation of SIE impacted Gen Z's BI who earn low-income (H3a;  $\beta = 0.232$ , t-value = 5.071, p < 0.001), contrary, the high-income

level was rejected ( $H3b$ ;  $\beta = -0.013$ ,  $t$ -value =  $-0.147$ ,  $p = 0.883$ ). The higher-class individual made more choices that make them stand out from others, on the contrary, the choices of lower-class individuals were likely to rely more on the other's [27]. Moreover, the lower-class group has a high tendency of social engagement in which more likely to follow the social norms rather than the upper-class group [28].

In the same vein, FC directly reflect UB for both Gen Z who has a low-income ( $H4a$ ;  $\beta = 0.312$ ,  $t$ -value =  $5.107$ ,  $p < 0.001$ ) and high-income level ( $H4b$ ;  $\beta = 0.469$ ,  $t$ -value =  $4.337$ ,  $p < 0.001$ ), similarly, BI affecting the UB of both low-income ( $H5a$ ;  $\beta = 0.761$ ,  $t$ -value =  $12.045$ ,  $p < 0.001$ ) and high-income group ( $H5b$ ;  $\beta = 0.506$ ,  $t$ -value =  $5.761$ ,  $p < 0.001$ ). This finding consequences with some authors which supported the UB of low-income people in India affected by FC [29]. Lastly, there was a strong confirmation of BI reflecting UB and this result is in the same direction as the prior studies [29–31].

## 5. Implications

The policymakers, online transaction facility providers, as well as commercial banks are necessary to understand their targeted consumers to maintain their competitive advantages. Specifically, Gen Z is the most important group of consumers in which potentially drive future economies [5], hence, the aforesaid parties need to communicate the right message to them. The marketing communications for a low-income group should include friends, relatives, or important persons in the communication tools because their intention was affected by SIE. Besides, the key messages in the communication are important to show the benefits, advantages, and usefulness of payments and systems in which reflect the low-income group intention. On the other hand, EE is only a factor reflecting higher-income group intention, thus, the illustration of methods for using payments and systems are crucial to include in the communication as guidance. Moreover, user-friendly applications, systems, or services are essential to be established to eliminate any difficulties and misunderstanding when using. FC is the common factor that determined both low- and high-income group's behavior, therefore, the availability of necessary resources via online platforms would enhance their preferences. Also, user support could be available when the consumers facing any difficulties regarding payments, applications, or systems usage. Once those implications have been met, the targeted consumers would be satisfied and further influenced their peers or friends to use it.

## 6. Conclusion

The current study adopting the UTAUT model to examine the e-payment intention and usage of Gen Z during the time of the Covid-19 pandemic. To better understand the current situation, the income levels have been categorized into the low-income and high-income levels as the study groups. The significant result shows the high-income group is likely to use e-payment services more than the lower-income group following their income level. Moreover, based on the original UTAUT, performance expectancy and social influence have a significant reflection on Gen Z's intention who earn a low income. Also, facilitating conditions and intentions have a direct effect on the use behavior of the lower-income group. Conversely, we found that effort expectancy has not predicted their intention. In the view of Gen Z who earn a higher income, performance expectancy and

social influence expectancy have no significant prediction on the intention, contrary, effort expectancy is the only factor that has a significant effect on their intention. Besides, there was strong evidence of facilitating conditions and intention directly influence their use behavior. Finally, there are some limitations in this research paper. The sample group of this study is limited to generation Z in Thailand. A cross-generational cohorts' study is suggested. Narrowing the scope of the study in another's form of e-payment services, for instance, m-banking, m-payment, mobile wallets, etc. would show the significant difference in results. Besides, the other demographic moderators e.g., gender have been ignored which might be causing unidentified hidden moderators. Lastly, there might be other factors that affect the intention and usage during the Covid-19 crisis which are excluded in the research framework. Therefore, future research might further study the cross-generation cohorts, cross-e-payment systems, as well as extending the UTAUT model to capture the world's most recent issues and situations.

## Acknowledgment

This research was financially supported by Khon Kaen University International College (KKUIC). We thank our colleagues from Khon Kaen University International College and EXPEDITE / Global Management Consulting Center, Khon Kaen University who provided insight and expertise that greatly assisted the research.

## References

- [1] World Health Organization. ACT now, ACT together 2020-2021 impact report [Internet]. 2021. Available from: <https://www.who.int/publications/m/item/act-now-act-together-2020-2021-impact-report>
- [2] Morgan PJ, Long TQ. Impacts of COVID-19 on households in ASEAN countries and their implications for human capital development [Internet]. Tokyo; 2021. (ADBI Working Paper). Report No.: 1226. Available from: <https://www.adb.org/publications/impacts-covid-19-households-asean-countries>
- [3] United Nations Conference on Trade and Development. Covid-19 and e-commerce: a global review [Internet]. Geneva; 2021. Available from: [https://unctad.org/system/files/official-document/dtlstict2020d13\\_en.pdf](https://unctad.org/system/files/official-document/dtlstict2020d13_en.pdf)
- [4] Capgemini Research Institute. World payments report 2020 [Internet]. 2020. Available from: <https://worldpaymentsreport.com/resources/world-payments-report-2020/>
- [5] Oxford Economics. Gen Z's role in shaping the digital economy [Internet]. London, UK; 2021. Available from: <https://www.oxfordeconomics.com/recent-releases/Gen-Z-role-in-shaping-the-digital-economy>
- [6] Gomez K, Tiffany M, Kimberly B, Sapp K, Brown A, Santner K. Welcome to generation Z [Internet]. 2018. Available from: <https://www2.deloitte.com/us/en/pages/consumer-business/articles/understanding-generation-z-in-the-workplace.html>
- [7] Rangan VK, Chu M, Petkoski D. The globe: segmenting the base of the pyramid [Internet]. Harvard Business Review. 2011 [cited 2021 Aug 24]. Available from: <https://hbr.org/2011/06/the-globe-segmenting-the-base-of-the-pyramid>
- [8] United Nations Development Programme. Creating value for all: strategies for doing business with the poor [Internet]. New York; 2008. Available from: [file:///C:/Users/Tomoko/Downloads/RW\\_rp\\_Creating\\_Value\\_for\\_All\\_Doing\\_Business\\_with\\_the\\_Poor.pdf](file:///C:/Users/Tomoko/Downloads/RW_rp_Creating_Value_for_All_Doing_Business_with_the_Poor.pdf)
- [9] Yurdakul D, Atik D, Dholakia N. Redefining the bottom of the pyramid from a marketing perspective. *Mark Theory* [Internet]. 2017 Sep 11 [cited 2021 Aug 24];17(3):289–303. Available from: <https://journals.sagepub.com/doi/abs/10.1177/1470593117704265>
- [10] Ministry of Labour. New minimum wage rate table (no.10) [Internet]. 2020 [cited 2021 Jul 6]. Available from: <https://www.mol.go.th/en/minimum-wage/>
- [11] Venkatesh, Morris, Davis, Davis. User acceptance of information technology: toward a unified view. *MIS Q* [Internet]. 2003 Jan 6;27(3):425. Available from: <http://www.jstor.org/stable/30036540>
- [12] Alalwan AA, Dwivedi YK, Rana NP. Factors influencing adoption of mobile banking by Jordanian bank customers: extending UTAUT2 with trust. *Int J Inf Manage*. 2017 Jun 1;37(3):99–110.

- [13] Oliveira T, Faria M, Thomas MA, Popović A. Extending the understanding of mobile banking adoption: when UTAUT meets TTF and ITM. *Int J Inf Manage* [Internet]. 2014 Oct 1 [cited 2021 Jan 4];34(5):689–703. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0268401214000668>
- [14] Joshi A, Kale S, Chandel S, Pal D. Likert scale: explored and explained. *Br J Appl Sci Technol.* 2015;7(4):396–403.
- [15] Cochran WG. Sampling techniques third edition. 2007;75.
- [16] Bentler PM, Chou C-P. Practical issues in structural modeling. *Sociol Methods Res* [Internet]. 1987 Aug 30;16(1):78–117. Available from: <https://doi.org/10.1177/0049124187016001004>
- [17] Byrne BM. Structural equation modeling with Mplus [Internet]. 1st ed. New York: Routledge; 2013. Available from: <https://www.taylorfrancis.com/books/9780203807644>
- [18] West SG, Finch JF, Curran PJ. Structural equation models with nonnormal variables: problems and remedies. In: Structural equation modeling: Concepts, issues, and applications. Thousand Oaks, CA, US: Sage Publications, Inc; 1995. p. 56–75.
- [19] Anderson JC, Gerbing DW. Structural equation modeling in practice: a review and recommended two-step approach. *Psychol Bull* [Internet]. 1988 May;103(3):411–23. Available from: <http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-2909.103.3.411>
- [20] Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate data analysis: Pearson new international edition. 7th ed. Pearson Education Limited; 2013.
- [21] Fornell, C., & Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. *J Mark Res.* 1981;18(1):39–50.
- [22] Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. *J Acad Mark Sci* [Internet]. 2015 Jan 22;43(1):115–35. Available from: <http://link.springer.com/10.1007/s11747-014-0403-8>
- [23] Bollen KA. Structural equations with latent variables [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1989. Available from: <http://doi.wiley.com/10.1002/978118619179>
- [24] Segars AH. Assessing the unidimensionality of measurement: a paradigm and illustration within the context of information systems research. *Omega* [Internet]. 1997 Feb;25(1):107–21. Available from: <https://www.sciencedirect.com/science/article/pii/S0305048396000515>
- [25] Ullman JB, Bentler PM. Structural equation modeling. In: *Handbook of Psychology* [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2003. p. 607–34. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/0471264385.wei0224>
- [26] Buhi ER, Goodson P, Neilands TB. Structural equation modeling: a primer for health behavior researchers. *Am J Health Behav* [Internet]. 2007 Jan 1;31(1):74–85. Available from: <http://openurl.ingenta.com/content/xref?genre=article&issn=1087-3244&volume=31&issue=1&spage=74>
- [27] Stephens NM, Markus HR, Townsend SSM. Choice as an act of meaning: The case of social class. *J Pers Soc Psychol* [Internet]. 2007;93(5):814–30. Available from: <http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.93.5.814>
- [28] Côté S. How social class shapes thoughts and actions in organizations. *Res Organ Behav* [Internet]. 2011 Jan;31:43–71. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0191308511000050>
- [29] Bai shya K, Samalia HV. Extending unified theory of acceptance and use of technology with perceived monetary value for smartphone adoption at the bottom of the pyramid. *Int J Inf Manage* [Internet]. 2020 Apr;51(September 2019):102036. Available from: <https://doi.org/10.1016/j.ijinfomgt.2019.11.004>
- [30] Makanyenza C. Determinants of consumers' intention to adopt mobile banking services in Zimbabwe. *Int J Bank Mark* [Internet]. 2017 Sep 4;35(6):997–1017. Available from: <https://www.emerald.com/insight/content/doi/10.1108/IJBM-07-2016-0099/full/html>
- [31] Jaruwachirathanakul B, Fink D. Internet banking adoption strategies for a developing country: the case of Thailand. *Internet Res* [Internet]. 2005 Jul;15(3):295–311. Available from: <https://www.emerald.com/insight/content/doi/10.1108/10662240510602708/full/html>

# Traffic Congestion Model in India by Shock Wave Theory

Tsutomo TSUBOI<sup>a,1</sup>

<sup>a</sup>New Business Creative Division, Nagoya Electric Works Co. Ltd., Japan

**Abstract.** This manuscript is a series of traffic flow analysis in India and it describes traffic congestion model by using shockwave theory from fluid mechanism. This is unique study for emerging country India traffic flow analysis during more than one month in October 2020. In order to chaotic traffic flow analysis in India, author observes one moth traffic flow data from the traffic monitoring cameras in 26 locations in a city where it is one of major city Ahmedabad in Gujarat states of India. In terms of traffic congestion, it is sued occupancy parameter of traffic flow data which indicates road occupancy percentage by vehicles. By using shock wave theory, author defines two traffic congestion model which are “forwarding traffic congestion” model and “stacking traffic congestion” model. These models explain traffic congestion condition and it is able to provide hint for solving traffic congestion problem in India.

**Keywords.** Traffic congestion, shock wave, traffic flow

## 1. Introduction

The traffic congestion becomes issues especially in emerging countries such as India these days, which causes CO<sub>2</sub> emission under growing number of vehicles because of economic growth. This issue is not only in India but also there is potential in any emerging countries. However it is big challenge to understand its traffic congestion because there is not enough traffic data and analyze traffic congestion mechanism. It is only clear condition, just too much vehicles and not enough infrastructure improvement such as road construction, traffic management system installation.

In general, traffic flow analysis has been started in 1933 when Greenshields B.D. [1] introduced traffic flow theory from his observation at the 13th Annual Meeting of the Highway Research Board. This theory is widely used for any traffic flow analysis and it also established as traffic flow fundamental equation from fluid flow analysis [2]. Therefore fluid flow theory is valid for considering traffic flow at the fundamental analysis. Therefore, we take shock wave theory for Indian traffic congestion analysis in this study.

After Greenshields traffic flow theory has been introduced, many traffic flow analysis has used and its theoretical validation has been confirmed. This is useful for understanding traffic flow characteristics such as traffic density ( $k$ ) to average vehicle speed ( $v$ ) characteristics, traffic density ( $k$ ) to traffic volume ( $q$ ) character etc. From these

---

<sup>1</sup> Corresponding Author, Tsutomo Tsuboi, New Business Creative division, Nagoya Electric Works Co. Ltd., 29-1 Ama-shi, Aichi, 490-1294, Japan; E-mail: t\_tsuboi@nagoya-denki.co.jp.

traffic flow analysis is valid to understand traffic flow condition but it is difficult to know micro traffic characteristics such as traffic congestion. Therefore, in traffic flow analysis, there are several challenges to use other flow mechanism about traffic congestion. One of flow mechanism is shock wave theory from fluid flow theory. This shock wave theory is also useful for explaining about traffic congestion in orderly traffic flow such as at highway traffic because moving behavior of vehicles is almost liner relationship between traffic speed and traffic volume. Therefore, it is challenge for make analysis at urban traffic flow, especially under developing countries such as India. The traffic condition in India as we know, is very chaotic and un-orderly driving behavior.

In the next section 2, it describes traffic flow observation location and environment. And the shock wave theory and occupancy parameter in the section 3. In the section 4, real traffic congestion analysis result is explained. In the section 5, this study summary and conclusion are described.

## **2. Observation environment,**

This study is joint research project between Japan and India as government funded for five years since April 2017. This project focuses on Low Carbon Society development for transportation in major cities of India.

### *2.1. Field*

The field is Ahmedabad city in Gujarat states of India and it locates in the west part of India. Its population is about 6 million at 2011. The current population is over 8 million [3] in 2018 from 5 million in 2011 and the number of vehicles is about 4 million in 2017 [4]. More than 70% vehicle is two wheelers, which is typical percentage in developing countries.

The official city profile of Ahmedabad is shown in Table 1.

**Table 1.** The city profile of Ahmedabad from local government “Ahmedabad Municipal Corporation or AMC.

Co-ordinates:	23.03° N 72.58° E
Area:	466 Sq.km. (year 2006)
Population:	55,77,940 (year 2011 Census)
Density:	11,948 /sq.km
Literacy Rate:	89.60 %
Average Annual Rainfall:	782 mm
Popularly known as:	Amdavad
STD Code:	079

## 2.2. Observation Environment

In this study, we have 31 traffic monitoring cameras in the city and its traffic data is collected at every minutes 24 hours per day. The camera's installation points are shown in Figure 1.

The camera's installation points are shown in Figure 1. The number shows each traffic monitoring camera location. Two digit number cameras are located in west side of block along with "132 Feet Ring Road" where there are many office, shops and market. The four digit number cameras are located near the river "Sabarmti River" so called River Front and new development area.

The River Front is control area by city government and it is not allowed to enter by auto-rickshaw and city bus. This road is currently used for a way to airport from city center. The main study location is on these both area because it is highly developing for business and the traffic grows heavily year by year. And there is another aspect which is metro development plan. The blue and red line in the map are Metro line which are under development by 2023.



**Figure 1.** Observation environment and location of traffic monitoring cameras.

In traffic flow analysis, there are two types of research, one is micro traffic flow analysis for example some major intersection traffic flow such as the area of camera #2001 through camera #2004—Paldi junction [5]. The other one is to focus on over all traffic condition analysis such as overall city traffic condition analysis [6]. In this study, we focus on traffic congestion not only micro area but also macro view. Especially macro view traffic congestion is important to know because traffic congestion is caused by several geographical environmental reason.

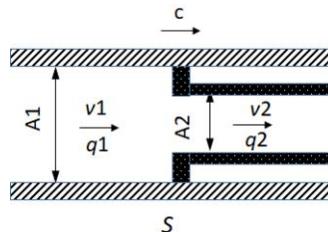
### 3. Traffic Flow Theory

In this section, we explain two traffic theory. One is shock wave and the other ones is occupancy parameter as traffic congestion condition parameter.

#### 3.1. Shock Wave

The shock wave theory comes from fluid mechanism originally but it becomes common for traffic flow analysis [7].

When there are two phases condition of a road which is illustrated in Figure 2, the number of vehicle passing boundary S is same between cross section A1 and cross section A2. When its number of vehicle at boundary S is  $N$ , Equation (1) is established. The road model in Figure 2 represents typical traffic congestion situation where driving vehicles go from wide area A1 to narrow space A2. The narrow point A2 is no necessary physically narrow because one traffic congestion occurs, front vehicles do not go smoothly, which is equivalent with physically narrow road.



**Figure 2.** Traffic Flow Shock Wave model.

$$N = (v_1 - c)k_1 t = (v_2 - c)k_2 t \quad (1)$$

where ( $c$ ) is moving speed at boundary S, ( $k$ ) is traffic density at each area 1 and 2, ( $t$ ) is time, and ( $v$ ) is vehicle speed at each area.

From Equation (1), Equation (2) is established.

$$(v_1 - c)k_1 = (v_2 - c)k_2 \quad (2)$$

Therefore, Equation (3) is established.

$$c = \frac{v_2 k_2 - v_1 k_1}{k_2 - k_1} = \frac{q_2 - q_1}{k_2 - k_1} \quad (3)$$

When  $q_2 - q_1 = \Delta q$  and  $k_2 - k_1 = \Delta k$ , Equation (3) is shown as Equation (4) by small changes of each parameters.

$$c = \frac{\Delta q}{\Delta k} = \frac{dq}{dk} \quad (4)$$

When it takes the differentiation by ( $k$ ) of both side, Equation (5) is established because of  $q = kv$ .

$$c = \frac{dq}{dk} = v + k \frac{dv}{dk} \quad (5)$$

From Greenshields equation,  $dv/dk < 0$  is established. Then Equation (6) is established.

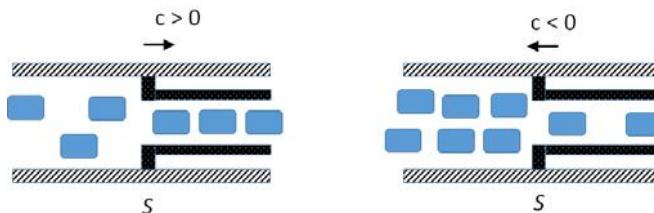
$$c < v \quad (6)$$

Therefore shockwave is transferred from the preceding vehicle in traffic flow to the subsequent vehicle. From traffic flow theory, critical traffic density is  $k_c$  and jam traffic density is  $k_j$ , Equation (7) is established and shock-wave is transferred forward direction and backward direction.

$$\left. \begin{array}{ll} c > 0 & \text{where } 0 \leq k < k_c \\ c = 0 & \text{where } k = k_c \\ c < 0 & \text{where } k_c \leq k < k_j \end{array} \right\} \quad (7)$$

From Equation (7), positive value ( $c > 0$ ) of shock wave means vehicle in front of section S becomes bottleneck by its congestion because the vehicle before section S has relative space between vehicles. On the other hand, negative value ( $c < 0$ ) of shock wave means vehicles before section S becomes bottleneck by congestion because the vehicle after section S has relative space between vehicles. Figure 3 shows this condition as illustration of traffic congestion model using the model of Figure 2.

When shock wave value equals to 0, there are two meanings—traffic has no congestion and moves smooth or traffic has completely congested and no movement.



**Figure 3.** Traffic Congestion model

Therefore,  $c > 0$  condition is defined “forward moving traffic congestion” and  $c < 0$  condition is “stacking traffic congestion”. From our experience, when traffic congestion starts from “forward moving congestion” to “stacking traffic congestion”.

### 3.2. Occupancy

In terms of traffic congestion parameter from traffic flow, occupancy parameter is used in general. The occupancy (OC) is defined the percentage ratio between total measurement time ( $T$ ) of the vehicles to a certain block of road section under certain moment. The formula is shown in Equation (6)

$$OC = \frac{1}{T} \sum_i t_i \times 100 (\%) \quad (6)$$

where  $T$  is time of measurement,  $t_i$  is detected time of vehicle  $i$  [8].

When number of existing vehicle a certain section is  $N$ , average length of vehicle is  $\bar{l}$ , Equation (7) is given.

$$OC = 100 \frac{q}{v} \bar{l} = 100 k \bar{l} \quad (7)$$

Therefore occupancy (OC) is proportional to traffic density ( $k$ ) and traffic volume ( $q$ ). From our previous study, occupancy shows actual traffic congestion condition [9]. And traffic congestion condition is normally 25% value.

### 3.3. Usage of traffic theory

From the previous traffic flow theory—shock wave and occupancy, the occupancy parameter indicates traffic congestion condition in general. Therefore, occupancy is the first choice parameter for judging traffic congestion condition from its traffic data. After identifying traffic congestion point in time basis and location basis, the next step focuses on traffic congestion patterns by using shock wave mechanism. In the previous section 2.1, there are two types of shock wave condition and they explain about traffic congestion model—forward forming model and backward recovery model. The forward forming shock wave shows the first stage of traffic congestion condition when its traffic congestion starts at the road. And the backward recovery shock wave shows heavy traffic congestion condition and vehicles which approach to the congestion area becomes stack situation because the front vehicles are not able to move because of its traffic congestion.

This traffic congestion analysis process is main research contexts in this study.

## 4. Discussion

Based on traffic measurement of the traffic monitoring cameras in Figure 1, Table 2 shows all occupancy value which is equal and greater than 25% condition. From October 2020 measurement data, the congestion occurs from 18:00 to 20:00 time frame.

**Table 2.** The occupancy measurement data in October 2020 at each location

	Time	18	19	20	21	22
Camera	2	33	59	47	37	26
	3	12	18	15	12	9
	4	19	29	24	20	15
	5	16	18	14	11	9
	6	3	4	4	5	5

7	15	20	15	12	10
8	13	14	9	10	6
9	20	25	21	18	17
10	11	13	9	8	6
11	24	35	25	19	15
13	15	21	16	12	9
1001	14	18	14	12	10
1002	12	15	11	9	8
1003	18	21	16	12	9
1004	6	8	6	6	5
1006	11	16	10	7	5
1007	9	10	7	5	5
1008	13	15	12	9	7
1009	8	11	8	7	5
1010	13	16	12	10	10
1011	11	13	10	8	8
1013	9	10	7	6	5
1014	7	9	6	5	5
1018	19	27	22	19	17
2003	14	22	19	16	14
2004	16	23	19	16	15

In Table 2, the highlighted occupancy value is equal and greater than 25%, which means its traffic condition is congested. There are five location in Figure 1. In Table 2, underline value is negative value of shockwave which is “stacking traffic condition”. In most case of traffic congestion occurs at 19:00 time frame. Based on this results, Figure 4 shows overall traffic density geographical image by ArcGIS tools. The value of Traffic density is vehicle/km·lane. We see the location camera #11, #2, #4 is most traffic congested area at 19:00 and most of location is “forward traffic congestion” condition.

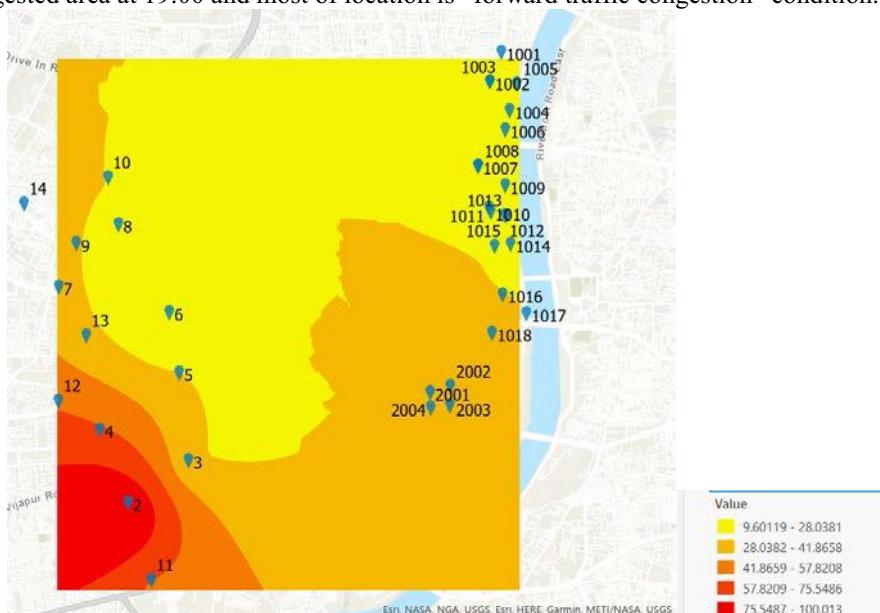


Figure 4. Traffic Density (vehicles/km·lane) in Ahmedabad at 19:00

## 5. Summary and Conclusion

In this study, we use shock wave analysis for Indian traffic congestion condition and it enable to identify the traffic congested time frame. In October 2020 traffic, the most traffic congested time is from 18:00 to 22:00 and particularly 19:00 is most crowded time. And the traffic congestion model is “forwarding traffic congestion” which means traffic congestion starts. On the other hand, “stacking traffic congestion” condition is very dedicated which happened at Camera #11 at 19:00 and Camera #2 at 22:00.

From this study, it is able to define clear traffic congestion time frame and location. It is necessary to study the reason why 18:00 to 22:00 is traffic congested time zone and the congested location is almost same area in Ahmedabad where is Camera #2 location. The area Camera #2 is commercial zone and there is big market in this area. The detail investigation of traffic congestion is future work.

As for conclusion, we show traffic congestion analysis in India by using traffic theory two parameters which is occupancy parameter for the indication about traffic congestion condition and shock wave parameter for the modeling traffic congestion condition. Those two parameter are useful to define traffic congestion condition for analyzing chaotic Indian traffic congestion. This method brings Indian traffic congestion mechanism how its traffic congestion occurs. It is necessary to analyze more area of Indian traffic data such as other month time period of traffic data in the next study. And it is also necessary to check the actual traffic congestion condition to visit on-site observation especially most congested point and time frame which we have already had some result of this study.

## Acknowledgement

This research is part of SATREPS program 2017 (ID: JPMJSA160 between India and Japan

## References

- [1] Kuhne R.D, Foundations of Traffic Flow Theory I: Greenshields' Legacy – Highway Traffic. Available from [https://www.krbalek.cz/For\\_students/mds/clanky/Greenshields.pdf](https://www.krbalek.cz/For_students/mds/clanky/Greenshields.pdf), [Accessed: 2021 Sep.]
- [2] Greenshields B.D., A Study of Traffic Capacity, Proc.H.R.B, 1935 , 14, pp.448-477.
- [3] Population of India [Internet] 2020 Available From: <https://indiapopulation2019.com/population-of-ahmedabad-2019.html> [Accessed: 2020-08-21]
- [4] Registered number of vehicles Ahmedabad India FY 2006-2017 Available From: <https://www.statista.com/statistics/665754/total-number-of-vehicles-in-ahmedabad-india>. [Accessed: 2020-08-21]
- [5] Tsuboi T., Traffic Congestion “Gap” Analysis in India, 7th International Conference on Vehicle Technology and Intelligent Transport Systems, 2021
- [6] Tsuboi T, Visualization and Analysis of Traffic Flow and Congestion in India, Infrastructures (ISSN 2412-3811) on 01 March 2021
- [7] Gupta A., Katiyar V.K., Analyses of shock waves and jams in traffic flow, Journal of Physics A Mathematical and General 38(19):4069, April 2005.
- [8] Kawakami.S, Matsui.H, Transport Engineering, Morikita Publishing Co., 2007, pp.102–103
- [9] Tsuboi T., Traffic Congestion Visualization by Traffic Parameters in India, The 2nd International Conference on Intelligent Control and Computing (ICICC) 2019.

# Detection for All Zero Coefficient Blocks in HEVC Based on Uniform Quantizer

Nana SHAN<sup>a,1</sup>, Henglu WEI<sup>b</sup>, Wei ZHOU<sup>c</sup>, and Zhemin DUAN<sup>c</sup>

<sup>a</sup>Taishan University, China

<sup>b</sup>Tsinghua University, China

<sup>c</sup>Northwestern Polytechnical University, China

**Abstract.** Transform and quantization are adopted in HEVC. There are lots of all zero coefficient blocks in transform and quantization. By detecting all zero coefficient blocks, the complexity of transform or quantization can be greatly reduced. All zero coefficient blocks for uniform quantizer can be efficiently detected by comparing the float quantization level of the estimated coefficients with an explicit threshold. The experimental result shows that 50% complexity of transform or quantization for uniform quantizer can be reduced with negligible loss of video coding efficiency.

**Keywords.** HEVC, all zero coefficient blocks, uniform quantization, RDO quantization, SATD

## 1. Introduction

High Efficiency Video Coding (HEVC) is the new video coding standard developed by JCT-VC [1]. HEVC provides better compression performance than H.264/AVC [2]. Several new techniques are included in HEVC, such as Rate Distortion Optimized Quantization (RDOQ), Sample Adaptive Offset (SAO), Coding Tree Units (CTU).

Transform and quantization are the most computational parts in HEVC [3]. The basic unit for transform and quantization in HEVC is called transform block (TB). Skipping all zero coefficient blocks can reduce the coding complexity of transform and quantization, as well as the prediction mode decision and advanced motion vector prediction [4].

There are many proposed detection algorithms for all zero coefficient blocks in H.264/AVC. An early predicting all zero coefficients is proposed in [5], and the threshold can be changed with the quantization level. A hybrid model was proposed in [6] to predict zero quantized Discrete Cosine Transform (DCT) coefficients with Gaussian distribution and optimized efficient condition to detect all zero DCT blocks. To get more frequency characteristics, Xie presents a mathematic model by analysing DCT for detecting all zero coefficient blocks [7].

The above algorithms mainly focused on detecting all zero coefficient blocks by mathematic models without considering uniform quantizer (UQ). In this paper, the

---

<sup>1</sup>Corresponding Author, Nana Shan; School of Information Science and Technology, Taishan University. E-mail: helensnn@hotmail.com. This work was supported by the Scientific Research Foundation of Taishan University (Grant No. Y-01-2017003).

detecting all zero coefficient blocks algorithm for UQ is proposed. And the proposed algorithm is based on determining the quantization level (QL) for the global maximum magnitude (MM). The estimated global MM is used to detect all zero coefficient blocks (AZCB) and non-all zero coefficient blocks (non-AZCB). As a result, AZCB can be easily detected by comparing the float QL (FQL) of the estimated coefficients with the threshold.

## 2. Transform and Quantization in HEVC

The HEVC standard specifies integer DCT matrices of size  $4\times 4$ ,  $8\times 8$ ,  $16\times 16$  and  $32\times 32$  to be used for two-dimensional transforms in the context of block-based motion-compensated video compression. For the  $4\times 4$  transform of luma intra picture prediction residuals, an integer discrete sine transform (DST) is alternatively specified. In addition, transform skip is also introduced in HEVC to improve compression ratio of screen-content video sequences generated in applications such as remote desktop, slideshows etc. When transform skip mode is used, prediction residuals are quantized directly. Transform skip mode is restricted to only  $4\times 4$  transform blocks. DST is regarded as a special DCT, and transform skip is regarded as a special transform.

By supposing the transform of HEVC encoder is orthonormal in this paper, UQ can be calculated as,

$$l_i^{uq} = \text{floor} \left( \frac{|c_i|}{Q_{step}} + \text{offset} \right) \quad (1)$$

where  $c_i$  is transform coefficient,  $l_i^{uq}$  is the QL of  $c_i$ , and  $Q_{step}$  is the step of quantization. The offset depends on the slice type, for I slices the offset equals  $1/3$  and for non-I slices it is  $1/6$ .

## 3. DCT Coefficients Estimation

Zero mean Gaussian distribution model is used in this paper to get the distribution of prediction residuals. Then, the distribution of DCT coefficients can be calculated. The standard deviation of DCT coefficients are used to estimate the global MM.

DCT coefficients are estimated twice in the original method. The proposed method use transform coefficients estimation based on the feature of DCT energy concentration. As Figure 1 shows, in the DCT process, prediction residuals are concentrated with more low frequency coefficients, low frequency coefficients have higher impacts to the coding units than high frequency coefficients. So, we should focus on high frequency coefficients.

There is a threshold of high frequency coefficients and low frequency coefficients. From the classification result shows in Figure 1, coefficients in  $4\times 4$  TB are all selected as low frequency coefficients. But for  $8\times 8$ ,  $16\times 16$  and  $32\times 32$  TBs, only part coefficients are selected as low frequency coefficients and most of them are gathered at the left-top of the coding units.

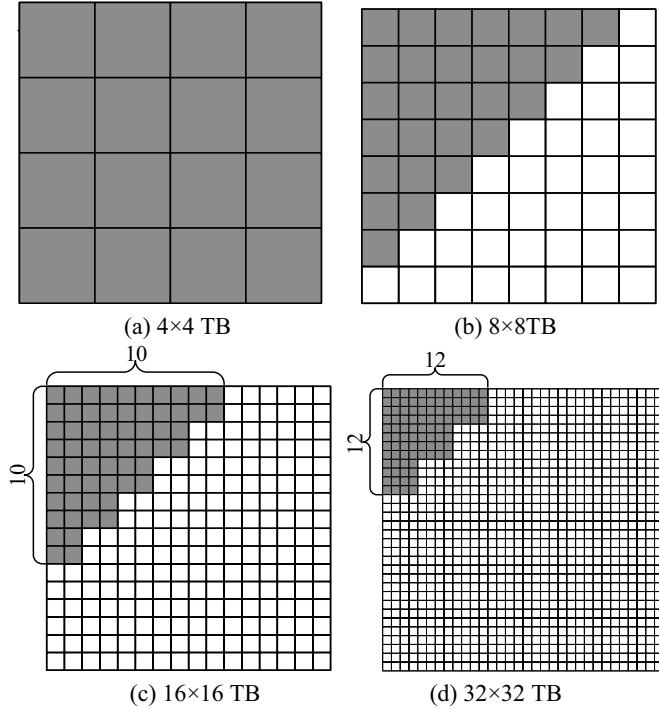


Figure 1. The selected low frequency coefficients

#### 4. AZCB detection in HEVC

The limit between QL equals zero and non-zero is explicit for UQ. And FQL can be defined as,

$$l_i^{float} = \frac{|c_i|}{Q_{step}} \quad (2)$$

Comparing (1) and (2), it can be found that  $l_i^{float}$  is just the float style of  $l_i^{uq}$ . What is more, if  $l_i^{float}$  is less than  $(1 - offset)$ ,  $l_i^{uq}$  equals to zero. Set  $Th\_U = (1 - offset)$ , then  $Th\_U$  can be used as the threshold between QL equals zero and non-zero. If  $Th\_U$  is larger than FQL of every coefficient in a TB, AZCB can be determined. Therefore, the FQL of the estimated low frequency coefficients and  $HM_{est}$  should be compared with  $Th\_U$  one by one until AZCB or non-AZCB can be determined. But it is found that the accuracy of AZCB detection is sensitive to the MM of the estimated low frequency coefficients. The estimated low frequency coefficients (HT coefficients) are amplified a bit before AZCB detection to prevent the misclassification of the non-AZCB.

The scaling factor is calculated as,

$$\alpha_N = \text{mean}_{u,v \in LE}(\sigma_{D/H}(u,v)), \sigma_{D/H}(u,v) > 1 \quad (3)$$

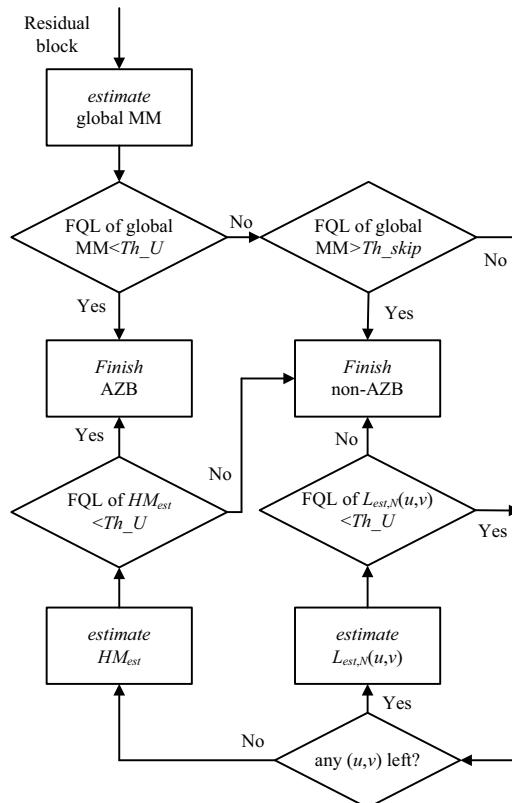
where  $LF$  is the location set of the selected low frequency coefficients,  $N$  is the size of TB and  $\sigma_{D/H} = \sigma_D / \sigma_H$ .  $\sigma_D$  and  $\sigma_H$  are the standard deviation of DCT coefficients and the standard deviation of HT coefficients respectively. These two parameters can be

obtained by replacing  $D$  by DCT or HT. Then, the estimated low frequency DCT coefficient is,

$$L_{est,N}(u,v) = \alpha_N \cdot B_N(u,v), \quad u,v \in LF \quad (4)$$

Besides, the global MM can also be used to pre-detect AZCB. It is obvious that if the QL of the global MM is zero, the QLs of other coefficients must be zero. Therefore, AZCB can be easily determined by comparing the FQL of the global MM with  $Th\_U$ . If the FQL of the global MM is less than  $Th\_U$ , AZCB can be determined. The estimated global MM is also used to skip coefficients estimation when non-AZCB occurs. The threshold in this paper is assigned to be 1.3 for SAD and 1.9 for SATD. If the FQL of the estimated global MM is larger than the defined threshold, denoted by  $Th\_skip$ , the TB is considered as a non-AZCB, and coefficients estimation will be skipped.

The flowchart of AZCB detection in UQ is shown in Figure 2. At first, the global MM of DCT coefficients is estimated and its FQL is compared with  $Th\_U$ . If  $Th\_U$  is larger than the FQL, AZCB is determined; if not the FQL will be compared with  $Th\_skip$  to determine non-AZCB. If AZCB or non-AZCB cannot be determined, then low frequency coefficient  $L_{est,N}(u,v)$  is estimated and the corresponding FQL is compared with  $Th\_U$  one by one to determine non-AZCB. If non-AZCB cannot be determined by low frequency coefficients, the MM of high frequency coefficients  $HM_{est}$  will be calculated and its FQL is compared with  $Th\_U$ . If  $Th\_U$  is larger than the FQL, the corresponding TB is an AZCB, if not it is a non-AZCB.



**Figure 2.** The flowchart of AZCB detection in UQ

## 5. Experimental results

The following experimental results are based on HEVC test model version 13.0 (HM 13.0). The standard encoder\_lowdelay\_main.cfg configuration file is used. 21 typical test sequences are selected from Class A to Class F, with the quantization parameters (QP) set to 22, 27, 32, and 37 respectively. The change of BDPSNR, BDBR [8], and coding time are shown on Table 1. As Table 1 shows the average time saving of the proposed method is 50%. The proposed method can greatly reduce coding efficiency with a negligible loss of PSNR and BD-rate.

**Table 1.** The Results for UQ AZCB Detection

Sequence	BDBR (%)	BDPSNR (dB)	TS (%)
PeopleOnStreet	-0.04	0.002	44
Traffic	-0.15	0.005	50
BasketballDrive	0.09	0.000	44
BQTerrace	-0.07	0.001	51
Cactus	-0.09	0.002	48
Kimono	-0.07	0.002	42
ParkScene	0.03	-0.001	53
BasketballDrill	0.09	-0.004	49
BQMall	-0.06	0.003	44
PartyScene	-0.03	0.002	41
RaceHorsesC	-0.05	0.002	36
BasketballPass	0.00	-0.001	43
BlowingBubbles	0.14	-0.005	41
BQSquare	0.13	-0.005	50
RaceHorsesD	0.02	-0.002	36
FourPeople	0.11	-0.006	64
Johnny	-0.04	-0.001	69
KristenAndSara	0.12	-0.003	65
ChinaSpeed	-0.09	0.005	51
SlideEditing	-0.07	0.010	69
SlideShow	0.10	-0.007	63
Average	0.00	0.000	50

The accuracy of AZCB detection process can be measured by false acceptance ration (FAR) and false rejection ratio (FRR). They can be calculated as equation (5),

$$\begin{aligned} FRR &= \frac{\text{num}_{mz}}{\text{num}_z} \\ FAR &= \frac{\text{num}_{mn}}{\text{num}_n} \end{aligned} \quad (5)$$

The  $\text{num}_{mz}$  means the number of AZCB which are classified as non-AZCB incorrectly, and  $\text{num}_z$  represents the real number of AZCB. While  $\text{num}_{mn}$  and  $\text{num}_n$  represent the number of non-AZCB which is misclassified as AZCB and the real number of non-AZCB. FRR means the ratio of AZCB which are classified as non-AZCB incorrectly, while FAR means the ratio of non-AZCB misclassified as AZCB. We want to detect AZCB as much as possible, which means that the FRR and FAR will be very low. In other words, higher detection efficiency means lower FRR and FAR.

**Table 2.** Detection Efficiency in UQ by the Proposed Algorithm(FRR)

Sequence	4×4 FRR (%)	8×8 FRR (%)	16×16 FRR (%)	32×32 FRR (%)
Class A	1.2	3.2	5.2	13.7
Class B	1.8	3.5	5.5	12.2
Class C	2.8	6.2	9.2	19.0
Class D	2.1	7.1	10.4	20.5
Class E	1.0	2.4	3.4	7.2
Class F	1.1	3.0	4.4	7.0
Average	1.7	4.2	6.4	13.3

The results of the proposed algorithm for UQ are shown as Table2 and Table3. The average FRR of 4×4, 8×8, 16×16 and 32×32 TBs are 1.7%, 4.2%, 6.4% and 13.3% respectively. The average FRR of 4×4, 8×8, 16×16 and 32×32 TBs are 2.9%, 4.1%, 6.8% and 5.8% respectively.

**Table 3.** Detection Efficiency in UQ by the Proposed Algorithm(FAR)

Sequence	4×4 FAR (%)	8×8 FAR (%)	16×16 FAR (%)	32×32 FAR (%)
Class A	3.0	4.3	6.5	4.0
Class B	3.0	4.4	7.0	6.3
Class C	2.4	3.4	5.0	3.4
Class D	3.2	4.4	6.1	5.0
Class E	3.2	5.2	10.5	10.8
Class F	2.3	3.1	6.2	5.7
Average	2.9	4.1	6.8	5.8

Figure 3 compared the frame for both the proposed AZCB algorithm and the original HM test model. There are no notable differences between these images.

**Figure 3.** frame of proposed algorithm and HM

## 6. Conclusion

An AZCB detection algorithm for UQ is proposed in this paper. The algorithm adopts a low and high frequency coefficients separation strategy to estimate transform coefficients. Threshold to determine AZCB in UQ can be deduced directly. The proposed algorithms are implemented on HEVC test model to verify the efficiency. Experiment results show that the proposed algorithm can efficiently reduce the computation complexity while keeping nearly the same RD performance as with the original algorithm in HEVC.

## References

- [1] International Telecommunication Union Telecommunication Standardization Sector (ITU-T) and International Organization for Standardization/International Electrotechnical Commission (ISO/IEC), "High efficiency video coding." Rec. H.265/ISO/IEC 23008-2, April 2013.
- [2] Sullivan, G. J., Ohm, J., Han, W. J., & Wiegand, T. Overview of the high efficiency video coding (hevc) standard. *IEEE Transactions on Circuits & Systems for Video Technology*, 22(12), 2012, 1649-1668.
- [3] Jung, S. W., Baek, S. J., Park, C. S., & Ko, S. J. Fast mode decision using all-zero block detection for fidelity and spatial scalable video coding. *IEEE Transactions on Circuits & Systems for Video Technology*, 20(2), 2010, 201-206.
- [4] Ou, J. H. P. L. S. (2003). Motion search method based on all zero block detection in h.264. *Signal Processing*.
- [5] Weicai Z, Jing L , Lichcng J , et al. A novel early prediction method of all-zero coefficients in h.263. *journal of electronics&information technology*, 2003, 25(4):573-576.
- [6] Wang, H., & Kwong, S. Hybrid model to detect zero quantized dct coefficients in h.264. *Multimedia IEEE Transactions on*, 9(4), 2007, 728 - 735.
- [7] Xie, Z., Liu, Y., Liu, J., & Yang, T. A general method for detecting all-zero blocks prior to dct and quantization. *IEEE Transactions on Circuits & Systems for Video Technology*, 17(2),2007, 237-241.
- [8] Bjontegaard, G. Calculation of average psnr difference between rd-curves. ITU-T VCEG-M33, April, 2001.

# Studies on Platform-Type Organization Design Based on Leadership Promotion

Xin MA<sup>1</sup>, Shuhui LI, Wei YANG, Wenhao YAN, Aisheng LIU, and Yanting TAN

*Customer Service Center of the State Grid Corporation of China*

**Abstract.** In the present scenario, continuous reforms in the social organization have become a new normal of maintaining the survival and development of enterprises and enhancing the competitiveness on the market, wherein the core of revolution is the change of management mode and leadership promotion. In this pursuit, the present study envisaged the innovation of organizational management mode of a subordinate secondary department of the Customer Service Center of the State Grid Corporation of China. Meanwhile, the revolution from the traditional bureaucratic organization to a novel platform-typed organization for the state-owned enterprises in internet evolution process was analyzed via the promotion of cadre leadership based on dual-requirements. The characteristic organization form with dual identities was built to better support the business development and compliance management, thereby achieving greater economic and social value creation for the enterprises.

**Keywords.** state-owned enterprises; Internet-based; Internet evolution; leadership; revolution of organization

## 1. Introduction

The revolution in a company organization is generally triggered by the adjustment of development strategy programming and evolves with the variations in the market and technological development. During the evolution of state-owned enterprises from the traditional bureaucratic organization to the internet-based platform organization, the devolution of the administrative power from the senior leaders to lower levels and the enhancement of leaderships of mid- and low-level cadres plays a key role in the success of the revolution of the organizational form and even the development strategy. The Customer Service Center of the State Grid Corporation of China (hereinafter referred to as the Customer Service Center or the Center) performs reconfiguration of the management factors, breaks the originally ossified bureaucracy management system, and establishes the platform-type organizations, with the aim of contributing to the enhancement of organization performance, ensuring high-speed business development, fast adaptation to ever-changing market changes and fast response to the customer requirements.

---

<sup>1</sup>Corresponding Author, Xin MA, Customer Service Center of the State Grid Corporation of China; Email: 57169222@qq.com

## 2. Design of lean management idea

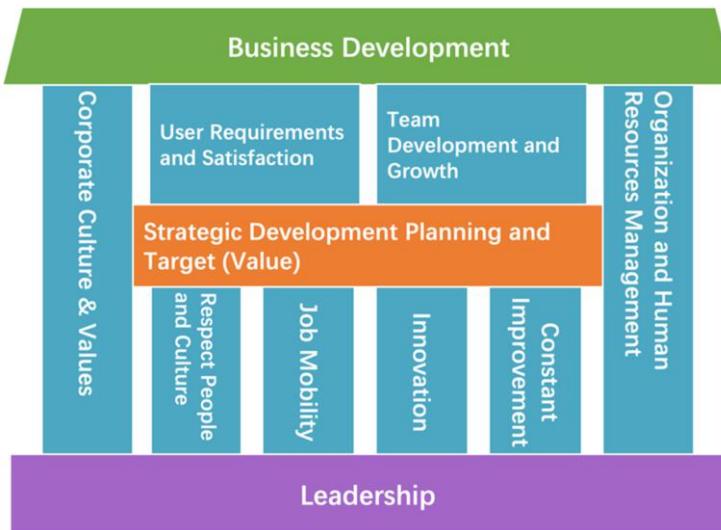
At present, the roots that shackle company development lie in the fact that the current organization structure cannot bear the internet-based development strategy, thereby failing in matching with the rapid iteration speed of the internet products. The employees, in particularly the introduced talents, face the contradiction between the traditional management requirements and the internet business development. They struggle to cope with the busy situation induced by mismatching between the job responsibilities and contents. Additionally, they are overburdened with work due to the stortage of staff and thereby cannot achieve their own values. On the other hand, the promotion channels are narrow, salary incentives are relatively weak and the personal development becomes impossible. In order to solve the contraction between the current rapid business development and the relatively weak organization, a task of top priority is to optimize both the organization and management system, and match and promote the productivity enhancement with more advanced productive relations. This will satisfy the requirements of the rapid business growth and achieve a sustained value creation.

In order to better adapt to the pulse of the market development and address a lot of problems, such as the mismatching between the internal management elements and the actual business requirements. The Customer Service Center of the State Grid Corporation of China, as a secondary administrative unit, investigated the internet enterprises such as Huawei[1] on the basis of originally traditional management mode of the state-owned enterprises. It also explored and merged into lean and agile management [2] thought of internet for guiding the overall design of the revolution. Moreover, under the background of supporting the construction of the new power system and achieving the combination of internet and customer service, the Customer Service Center broke the authoritarian leadership model, established the leadership rooted in high-, middle- and basic-level cadres, and builds a novel platform-typed organization (Figure 1 shows the house of leadership management ) that adopts the strategic goal of serving the customers and enhancing the customer satisfaction as core value pursuit. Based on the enterprise culture and values, the Customer Service Center laid an emphasis on the organization and human resource management, adopted communication as the link, respected employees' physical, intellectual, emotion security as well as their cultural requirements, and created a smooth working process along with a continuously improved working environment.

## 3. Revolution of the organizational structure

The most important thing in the core proposition of the organization development lies in the revolution of the organizational structure and the innovation in the talent mechanism. For the currently dispersive and variable market environment, the organizational structure should be flexible, so as to be compatible with the constant iteration and innovation of enterprises. This will become an important basis in the design of the platform-type organization.

In order to settle the problems like the mismatching between the organization mode and the business development degree, and the mismatching between the organization operating situation and rapidly iteratively market requirements, the secondary department explored the revolution in the organization structure on the basis of originally 4 departments and attempted the adjustments mainly in the following three aspects (Figure 2 shows the restructured organizational structure ).



**Figure 1.** house of leadership management

Firstly, the general management department was created to the supporting platform with shared functions. The Customer Service Center focused on the business support and services, and compliance supervision, and built the business partner (BP) group that satisfied the front business requirements. Additionally, it provided the front BPs with specific incentive schemes for upgradation of the configuration, thereby injecting different degrees of resources for different projects and ensuring benefits and full operation to the full extent. Meanwhile, the Customer Service Center provided the management on the projects with injected resources after approval and achieved the business knowledge energization (mainly including strategy, finance and human resource) for the front-end.

Secondly, the vertical business unit was transformed into the empowered mid-ground. Starting from the front-end requirements, the Customer Service Center integrates the precipitated resources and professional functions on the middle- and back-end, which could provide the modularized packing services and ability sinking services for the front-end organizations, thereby contributing to favorable performances on the front-end. In addition, performance management methods based on objective and key results (OKR) management was optimized by allocating enlightened and effective incentive policies and innovatively offering both inventive and energization mechanisms. The role of the mid-ground as the connector was released, so that the back-end resources could become the middlewares that could be called on demands, which energized both the professional resources on the front-end and professional knowledge with the delivery of middlewares. Meanwhile, the promotion channels for ordinary employees, technical experts, management and operation talents were built and both horizontal and vertical employee turnover mechanisms were set.

Thirdly, the front-end was elite. The Customer Service Center constructed three front-end business groups—ranger group, new business product group and traditional business product group. The customer requirements were scooped out continuously at multi-dimensions, so as to pull the organization by natural growth instincts at the front end. By reference to the organizational structure of agile development team, the product

groups were created, so that all staff followed the command of users. Accordingly, the back-end functional departments that originally did not come into contact with the users were reconstructed, and the organizational form in which the front end was responsible for interacting with users and understanding the users' rigid demands. The mid-ground and the back-end actively organized various types of resources on the platform, so as to continuously output the products, services or solutions in good consistence with user demands, achieve the integration between flexible product groups/teams and rigid departments, and ensure full empowerment on the front end by the leadership. Meanwhile, the self-regulation on the front end was strengthened, so as to ensure high-efficiency coordination among the various organizations on the front- and mid-grounds and ensure an effective establishment and operation of platform-typed organizations. Accordingly, the organization was not only a hierarchical organization, but also followed a moderately integrated and coherent-business organizational operating mode.

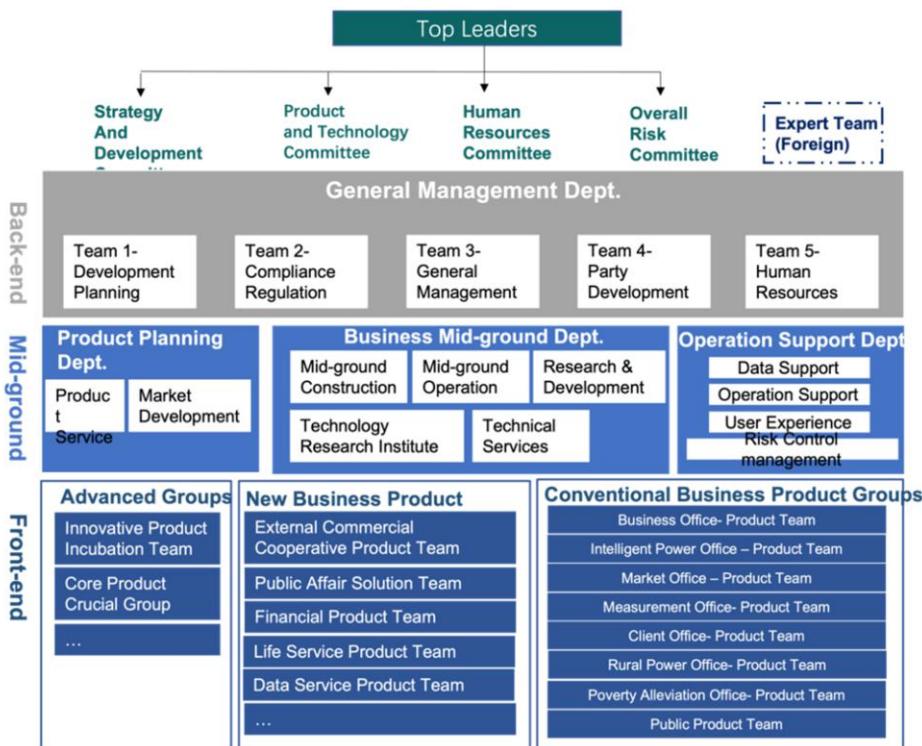


Figure 2. organizational structure diagram

During the overall organizational revolution process, the Customer Service Center obeyed the principle of small steps for running and continuous iteration, performed the adjustment of organization structures and gradually increased and adjusted the personnel. In accordance with the principle of inclination towards the front-line and healthy competition, the Customer Service Center gave the priority to the allocation towards the product groups, attempted resource allocation and dynamic adjustment within the product groups, and encouraged to preferentially allocate advantageous resources for advantageous products. In addition, based on the principle of talent pooling and

resource coordination, the outstanding employees with leadership were encouraged, who were allocated to new posts or product groups for training, so as to reserve high-quality versatile talents for organizations.

#### 4. Leadership promotion

An organization is composed of people. People are the important factor that achieves the organizational goal. A series of problems were encountered, including the mismatching of the current management mechanism and promotion channel with the personal development, professional posts and actual work requirements of personnels, the mismatching between the reserve of talents and business development, and the mismatching of performance management with the diversified positions and personnel management requirements. In order to solve these problems, the secondary department strived to achieve a synchronous revolution of the management and leadership, released the management functions of the middle- and basic-level leading cadres and broke the ceiling of enterprise development. By setting talent selection as the gripper that grasped the driving source of development, the department stimulated multi-level needs of people via talent selection, drove away the laziness in human nature and leveraged a greater value creation through reasonable distribution of rights.

The department mainly considered the revolution in the following three aspects: Firstly, the leader level clearly defined the keynote of self-targeted revolution, changed both the leadership concept and style, made the transition from command leadership to service leadership, promoted the fault-tolerant spirit, and supported and energized the team innovation. Meanwhile, the department constructed four-vertical, six-horizontal and three-qualified post rank sequences, helped the employees for career planning and guidance in accordance with their own characteristics and overall enhanced both the leadership and professional capabilities of employees. Accordingly, the organizational ability was effectively enhanced, so as to achieve win-win results of employee career development and enterprise management development. Secondly, the thought of agile values was publicized by the senior leaders, systematically arranged, dismantled and implemented by the mid- and basic-level leaders. By creating pilot product groups and forming benchmarking cases, rapid launching and validation of products was combined with material and immaterial incentives for team, thereby effectively stimulating the working and innovation enthusiasms of team. Thirdly, by reference to personnel capability evaluation and grading modes in internet companies, the key performance indicators (KPIs) were set in combination with objectives and key results (OKR) according to the department management requirements, and OKR indexes were set in combination with product innovation. Moreover, the corresponding performance assessment tool was designed, so as to assist strategic conduction, hierarchical classification and effective revolution of management mode, which provided a multi-dimensional development opportunities and growth space for talents. At the same time, the internal capital appreciation cycles featured by value creation, value evaluation and value distribution were re-constructed, while a just, fair and open healthy comparison, and a competition environment was built to achieve self-leading and self-driving of employees. This further contributed to the organizational self-operation.

## 5. Conclusion

In the present study, the secondary department under the administration of the Customer Service Center first analyzed and clarified the development strategic positioning and development mode in depth, and performed the research on platform-type organization revolution. By diagnosing and analyzing the problems in the implementation of matrix agile organizations in transverse process flow, the platform-type organizational mode featured by resource back-end, empowered mid-ground, custom front-end was investigated in combination with actual development and business requirements, and simultaneously, a novel organization management model and the supporting measures and mechanisms were constructed. After the selection and establishment of the institutional reforms and flexible teams, an organizational form abundant in novel leadership with diversified management and agile products was successfully built. The leading cadres at different levels could communicate and coordinate with employees, so that the development strategy could be decomposed and implemented step by step. The department also made innovations in the management and motivation modes, and encouraged the employees to work actively for achieving the organization's development goal. This ensured a good integration and mutual promotion between the management requirements of the state-owned enterprises and the development characteristics of internet.

## References

- [1] Tian Tao, David De Cremer and Wu Chunbo.,Huawei: Leadership, Culture and Connectivity.,New Delhi: SAGE Publications; 2018. 424p.
- [2] Dean Leffingwell. SAFe 4.0 Reference Guide: Scaled Agile Framework for Lean Software and Systems Engineering. China Machine Press; 2017. 356p.

# Interdisciplinary Framework: A Building Information Modeling Using Structural Equation Analysis in Lean Construction Project Management

Dante SILVA<sup>a,1</sup>, Kevin Lawrence DE JESUS<sup>a</sup>, Bernard VILLAVERDE<sup>a</sup>,  
Andrea Isabelle ENCISO<sup>a</sup>, Amanda Nicole MECIJA<sup>a</sup>, Justin Owen MENDOZA<sup>a</sup>

<sup>a</sup>School of Civil, Environmental, and Geological Engineering  
Mapúa University, Manila, Philippines

**Abstract.** The construction process and construction management are highly reliant on the interaction between the triple constraints of project management of quality: scope, time, and cost. The industry has incorporated certain principles and technology, such as lean-based construction principles and Building Information Modeling (BIM), to maximize the time, quality, and cost-efficiency of various construction projects. Analysis and assessment of the factors and functionalities are needed to show their synergic relationships and determine their significant impact on the construction project using Structural Equation Modeling (SEM). To gather information, a survey was conducted on the different construction companies in the Philippines. The results generated a model interrelating the triple constraints of project management and how they are affected by incorporating BIM and lean construction principles. The models created established a significant relationship towards all the triple constraints considering both individual and combined functionalities and factors. An interdisciplinary framework incorporating both BIM and lean principles is conformed to optimize construction stages based on the triple constraints of project management.

**Keywords.** Building Information Modeling (BIM), lean construction principles, triple constraints of project management, Structural Equation Modeling (SEM), interdisciplinary framework

## 1. Introduction

The construction industry has continually progressed from traditional to more advanced methods to achieve more high-quality outcomes for a shorter time and with fewer expenditures. With the increasing complexity of projects as time goes by, construction projects encounter a wide range of challenges that can either slow down the construction process or incur more costs such as natural calamities, public health emergencies, and countless unforeseen events. Construction management has different correlated parameters needed to consider, including the time, cost, and scope in preparation for the flow of the construction regardless of circumstances. Communication is an important factor in improving the productivity on site, therefore reducing problems. The

<sup>1</sup> Corresponding Author: Dante L. Silva, School of Civil, Environmental, and Geological Engineering, Mapúa University, Manila, Philippines, E-mail: dantesilva2000@yahoo.com

interdisciplinary factors associated with the issues in construction affects the progress of the external and internal site. [1]. Sustainability strategies take place to offer efficient and effective performances in the industry [2]. Two methodologies were implied to show the relationship of the triple constraints to the flow and process of the construction.

Building Information Modeling (BIM) is characterized as a digital model that imposes collaboration among project stakeholders [3]. The concept of BIM takes place through a life cycle of digital information, starting from conceptual design up to demolition. Its widespread use makes a significant effort in advancing the software and tools available in the market today [4,5]. On the other hand, lean construction was introduced to the construction industry from the principles being applied in the production industry from the Toyota Production System [6]. It provides continuous improvement in the industry by minimizing wastes to improve efficiency. These methods evolved throughout the years and are considered effective techniques applied for a strategic approach towards the construction sector [7].

The research is intended to formulate an interdisciplinary framework integrating both Building Information Modeling (BIM) and Lean Construction principles concerning the triple constraints of project management in the Philippine construction industry using Structural Equation Modeling (SEM). Provided their impact on output efficiency, the factors and functionalities of both BIM and Lean construction are integrated to determine the relationship of each parameter formed that affects the triple constraints throughout the different stages of construction projects.

## 2. Methodology

The methodology of the study is divided into five different phases. It involved researching online databases for preliminary research/information regarding the potential risks throughout the construction project phases. The designing of the survey questionnaire patterned from the existing information gathered on the related literature and studies. Cronbach's alpha (CA) was conducted to validate the survey questionnaire from the collected data from key construction companies in the Philippines. The researchers analyzed each component and identified the parameters needed to be focused on regarding the manifestation of Building Information Modeling (BIM) and lean construction principles alongside the triple constraints of project management: time, scope, and cost. The data being interpreted was evaluated to see its interaction through Structural Equation Modeling Analysis with measurement and structural model. The researchers defined the measurement model by utilizing Confirmatory Factor Analysis (CFA) procedures. Methods of factor computations were performed to alleviate the relationship of the variable with programming software, SPSS AMOS and Statistics. a complete measurement model was produced to emphasize the specified testing of hypotheses. Lastly, to finalize the full structured framework, the data collected from the measurement model is used for model fit assessment and path assessment and estimating squared multiple correlations. It is being interpreted to finally answer the seven (7) hypotheses set in this study in relation to the individual and combined triple constraints including time, cost, scope, time-scope, scope-cost, cost-time, and time-cost-scope and how it affects to the interaction of BIM and lean.

### **3. Data and Discussion**

#### *3.1. Preliminary Research on the Risks and Issues Encountered on Construction Projects*

The risks and problems encountered in construction are classified based on the people, process, and technology which organized the ideologies for streamlined flow. One common problem in the construction industry is a delay that may be caused by several factors including, but not limited to procurement, unforeseen events, low productivity rates, and change orders. These problems are encountered due to the lack of either time, budget, or scope.

#### *3.2. Designing the Survey Questionnaire*

The Cronbach alpha generated from the survey questionnaire was sufficient to satisfy the given conditions possessing acceptable values of 0.7 and above [8]. The reliability of the individual model with 24 observed factors (Cronbach  $\alpha = 0.77$ ) is higher than the combined model with 32 observed factors (Cronbach  $\alpha = 0.72$ ). Considering the whole assessment for both models supported, it was verified with a relatively good value of 0.83.

#### *3.3. Collection, Analysis, and Cleaning of Data*

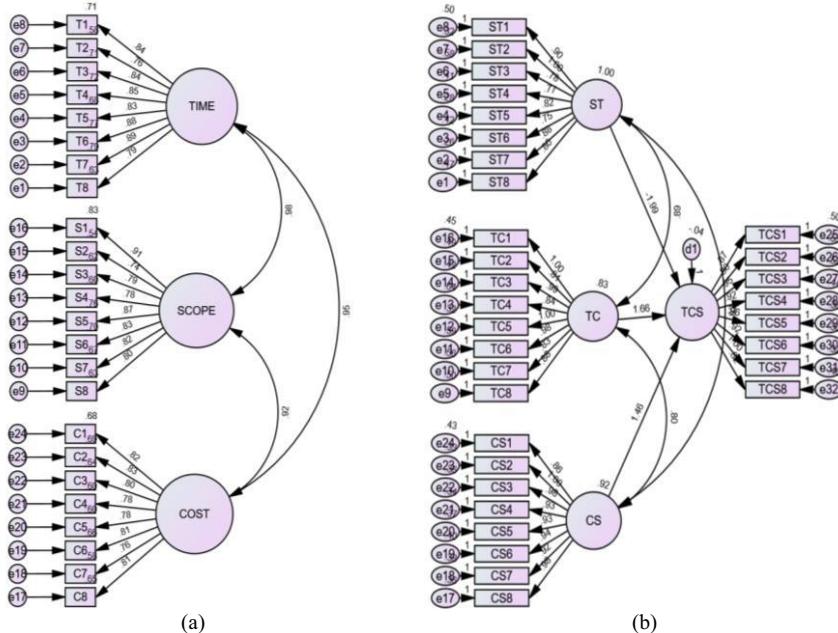
The researchers created hypothetical models that conceptualized the hypothesis tested in the study from the individual and combined factors. The triple constraints for both models were represented by ellipse or circle. Observed variables from the survey questionnaires were presented by rectangles. Single-head arrows were connected, representing the impact on another latent variable to the observed factors. Double-head arrows were used to show the covariances of the combined constraints. Errors of measurements are correlatedly associated with each observed variable, yet these measures are uncorrelated. Values and calculations needed for the analysis were generated using SPSS AMOS and Statistics.

#### *3.4. Evaluation through Confirmatory Factor Analysis*

In the analysis, considering the triple constraints of project management, the individual and combined factors satisfy the given 0.7 or higher factor loading criteria. Thus, all these factors show a positive and significant effect on the measured variables, supporting all hypotheses stated in the research study. Take note that the factor loadings for the latent variables in the prediction of the observed factors are significantly different from zero at the 0.001 level or two-tailed.

The researchers associated the variables by getting their Average Variance Extracted (AVE) and Composite Reliability (CR). The AVE measured the amount of variance that was implied based on the measurement errors. The CR is defined as the measurement of the internal consistency in scaling items. The ideal value of the Average Variance Extracted should be greater than 0.50, and the Composite Reliability should fall in the range of 0.70 or higher. [9]. The calculated AVE of all variables is found to be greater than 0.5 for both structured models, which indicates that all questions reflect the variables

indicated in the study. Moreover, the Composite Reliability of the three latent variables has a value greater than 0.7. This high value of CR indicates a high consistency among all the observed factors of the questionnaire. Provided both ideal values of the AVE and CR are achieved, the validity of the latent variables is established, and the model has good reliability.



**Figure 1.** Structural Model of (a) Individual Functionalities and Factors and (b) Combined Functionalities and Factors

Figure 1 shows the structural model for both the individual and combined functionalities, showing their factor loadings against one another. The factor loadings describe the correlation between two variables within the structural model with ideal values of 0.7 or higher. All observed factors that were considered for the model have significant relationships for the study since all values obtained were above the ideal value of 0.7. Overall, considering the factor loadings generated using the SPSS AMOS software, time and scope has the most significant relationship among all the variables, with 0.98 factor loading. Next is scope and time with 0.95, followed by cost and scope with 0.92 as their factor loadings.

### 3.5. Finalization of the Structured Frameworks

The researchers determined the discriminant validity through squared multiple correlations to assess the low correlation among the variables and identify distinctions among factors. To establish the discriminant validity, the value of the squared multiple correlations should be lower than the average variance extracted. For the discriminant validity for the individual functionalities, the variable of cost, the observed factors C1, C2, and C6 were not established. This is the same with the observed factors of S1, S5, S6, and S7 for the variable scope, and T4, T6, and T8. For the discriminant validity of the combined functionalities, it has two (2) latent variables with observed variables that

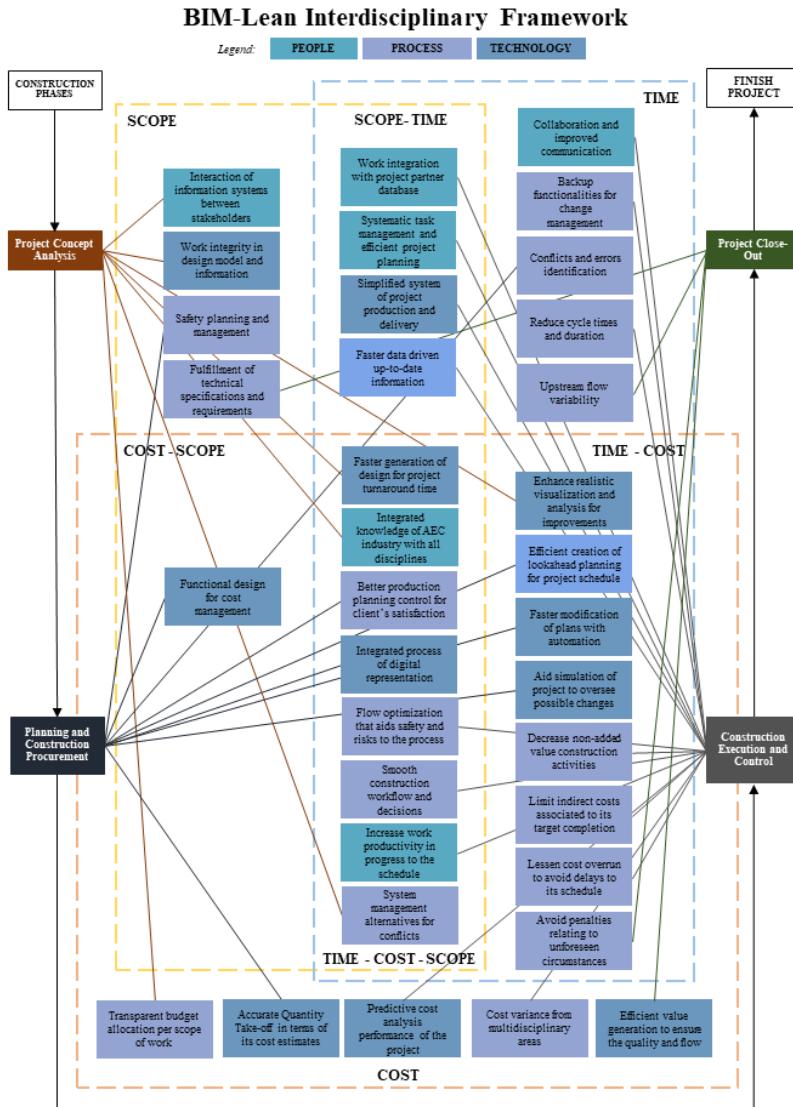
are not established: cost and scope (CS) and scope and time (ST). The majority of the observed variables for cost and scope were not established, apart from CS1. Half of the observed variables are not established in the latent variable of scope and time, namely ST1, ST2, ST5, and ST7. All observed variables under time and cost (TC) and time, cost, and scope (TCS) have a value of AVE higher than the squared correlation, which indicates an established discriminant validity. This information indicates that although the convergent validity is established and the observed factors significantly represent the latent factors, the observed ones not established correlate better with other latent variables than its parent variable.

Goodness-of-fit tests ensured that the data that were collected are normally distributed with eight (8) fit indices [10]. In both models, the fit indices generated using the data of this study showed acceptable values which indicates that the model can represent the entire population due to its normality.

**Table 1.** Goodness-Of-Fit Testing for Individual and Combined Models

Fit Indices	Parameter Estimates		Recommended Values
	Individual	Combined	
Chi-square ( $\chi^2$ ) P-value	Non-significant (p=0.069)	Non-Significant (p=0.053)	$\geq 0.05$
Normed Chi-Square ( $\chi^2 / df$ )	1.148	1.121	>3.00
Incremental Fit Index (IFI)	0.988	0.986	$\geq 0.90$
Tucker-Lewis Index (TLI)	0.983	0.981	$\geq 0.90$
Comparative Fit Index (CFI)	0.987	0.986	$\geq 0.90$
Root Mean Square Error of Approximation (RMSEA)	0.039	0.035	$\leq 0.05$
Goodness of Fit Index (GFI)	0.851	0.821	$\geq 0.80$
Adjusted Goodness of Fit Index (AGFI)	0.787	0.749	$\geq 0.90$

The interdisciplinary framework illustrated in Figure 2 summarizes the result of this research. Each latent variable were linked to the different stages of construction wherein these variables are related to. Consequently, each of the factor associated with divisions including the people, process, and technology. The initial roster of 56 latent variables were reduced to 35, eliminating those which are not established when the discriminant validly was tested. The observed variables included in the framework created was examined based on the assessed outcomes from their factor loadings. Out of the initial eight latent variables identified, only four was established with the scope, five with time, five with cost, four with scope-time, one with cost-scope, eight with time-cost, and eight with all three time-cost-scope. All the denied variables indicated that the following information does not established relationship towards the underlying data in connection to workflow reliability in construction [11]. Through this analysis, it can be determined that BIM and lean has its positive influence and correlations towards the triple constraints of project management showing that all the hypothesis satisfied and confirmed in the study. Furthermore, the results of varying path coefficients of SEM analysis suggested that the triple constraints have its great impact towards the project, supporting the claims on streamlined networks of communication, the dynamics and flexibility of triple constraints in quality project delivery, and optimized waste reduction [12,13].



**Figure 2. BIM-Lean Interdisciplinary Framework**

#### 4. Conclusion

The paper identified and tested the observed variables in relation to the factors and functionalities through Structural Equation Modeling Analysis. The research output of the structural model pointed out that usage of the Building Information Modeling and incorporation of Lean construction principles can result in the holistic improvement of the project scope, time efficiency, and a reduced overall cost to the project that can be a

basis to reinvent the risks and challenges in the construction industry. The application of BIM and lean significantly affects the construction project, starting from the project concept analysis up to the project closeout as the number of problems that arise throughout the project is significantly minimized. All the triple constraints of project management were interrelated with their individual and combined functionalities and factors in such way that it influences one another most especially in construction projects.

The interrelated factors with its focused approach are helpful towards the performance, reliability, and operations of the project. The main stakeholders such as the engineers, architects, contractors, and designers are influential professionals associated with integrating BIM and lean principles in terms of triple constraints as their collaborative efforts make significant relations to other stakeholders and clients involved in the project. Lastly, the effects of innovative technology such as birth of advanced software and hardware affect the triple constraints as well as possible risks and issues throughout the project duration.

## References

- [1] Silva DL, Macariola RN. Coping with the Information Age: Development of a Data Flow Diagram-Based Knowledge Management System for Mitigating Delays for Construction. OP Conf. Series: Materials Science and Engineering. 2019;652,12070.
- [2] Silva DL, Villaverde B, De Jesus KLM, Marcial Jr. L. Design Initiative Implementation Framework: A Model Integrating Kolmogorov-Smirnov in Sustainable Practices for Triple-Bottom-Line Principles in Construction Industry. 2020;8(4):1.
- [3] Schimanski CP, Marcher C, Monizza GP, Matt DT. The last planner system and building information modeling in construction execution: From an integrative review to a conceptual model for integration. Applied Sciences (Switzerland). 2020;10(3).
- [4] Borrmann A, König M, Koch C, Beetz J. Building Information Modeling: Technology Foundations and Industry Practice: Springer. 2018; p. 584.
- [5] Rodriguez LV, Bagcal OR, Baccay MA, Barbier BM. Adoption of Building Information Modeling (BIM) in the Philippines' AEC Industry: Prospects, Issues, and Challenges. ResearchGate. 2019 July; 9(2).
- [6] Kagioglou M, Koskela L, Tzortzopoulos P. Lean Construction Core Concepts and New Frontiers; 2020. 460
- [7] Parfenova E, Avilova ZN, Ganzha AN. Lean construction – an effective management system in the construction industry. IOP Conf Ser Mater Sci Eng. 2020; 945(1).
- [8] Taber KS. The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. Res Sci Educ. 2017 June; 48:1273–1296.
- [9] Ab Hamid MR, Sami W, Mohmad Sidek MH. Discriminant Validity Assessment: Use of Fornell & Larcker criterion versus HTMT Criterion. J Phys Conf Ser. 2017; 890(1): 012163.
- [10] Ismail NA, Idris NH, Ramli H, Muhammad Rooshdi RR, Sahamir SR. The relationship between cost estimates reliability and BIM Adoption: Sem analysis. IOP Conf Ser Earth Environ Sci. 2018; 117, 012045.
- [11] Zhang L, Chen X, Suo Y. Interrelationships among critical factors of work flow reliability in lean construction. J Constr Eng Manag. 2017 July. 23(5); 621-632.
- [12] Antunes R, Gonzalez V. A Production Model for Construction: A Theoretical Framework. Buildings 2015. 2015 Mar; 5(1): 209-228.
- [13] Heigermoser D, García de Soto B, Abbott ELS, Chua DKH. BIM-based last planner system tool for improving construction project management. Autom Constr. 2019 May; 104, 246-254.

# Mitigating Construction Deficiencies: An Impact Analysis for Low-Cost Housing Developments Utilizing Artificial Neural Network

Ramene U. LIM<sup>a,1</sup>, Dante L. SILVA<sup>b</sup>, and Kevin Lawrence M. DE JESUS<sup>a</sup>

<sup>a</sup>School of Graduate Studies, Mapua University, Manila, Philippines

<sup>b</sup>School of Civil, Environmental, & Geological Engineering, Mapua University, Manila, Philippines

**Abstract.** The aim of this study is to be able to come up with a supplemental project management policy guidelines and computational tool that will address the two major concerns in construction of low-cost housing, construction delays and workmanship defects. Through assessment of previous studies, factors causing delays and defects from the two major stakeholders involved in housing development projects were identified. With the use of the five-point Likert Scale in survey forms distributed to 60 professionals involved in housing development projects, factors were classified and identified according to its degree of impact on the overall construction efficiency. The statistics of these factors were organized and used to develop an Artificial Neural Network Model. The relative importance of the factors was measured using Garson's Algorithm. The derived equations from the developed ANN Model were used in formulating the computational tool and supplemental policy guidelines that can now be used to evaluate the workmanship defects and delay ratings of different housing developments. The computational tool was tested by 10 experts with their current projects and was able to receive a 4.6 out of 5 rubric evaluation rating, showing the tool's effectiveness in identifying and assessing the probability and impact of construction deficiencies on their projects.

**Keywords.** Artificial Neural Network, Construction Delays, Construction Management Program, Mass Housing Developments, Workmanship Defects

## 1. Introduction

Construction is one of the most consistent and growing industries in the whole world. This industry is one of the greatest economic drivers of countries. Here in the Philippines, one of the government's main goals is the "Build, Build, Build!" Project [1]. However, these major programs with regards to construction only focuses on large scale structures [1]. There are only few programs that address the housing needs of the Filipino people [2, 3]. There is an increase of 4% annually on the urbanization of major cities in the country [4, 5]. With this, urban settlers who have the financial capacity prefer high-rise condominiums, while others settle with rooms for rent. Horizontal housing developments

---

<sup>1</sup> Corresponding Author: Ramene U. Lim, School of Graduate Studies, Mapua University, Manila, Philippines; E-mail: ramenelim@yahoo.com

are left at the outskirts of the urban region and is handled with minimum supervision [6, 7], as it is an important aspect in construction management [8, 9]. This results in major workmanship defects and construction delays caused by the involved stakeholders. In addition, the world is currently facing a pandemic in Covid-19 that requires people to stay indoors due to the high transmissibility of this virus. This increased further increased the need for housing units for workers with work from home arrangements and for students taking online classes. Due to the increase in demand for housing units coupled with the current challenge in experiencing delays and defects in ongoing housing projects, it is important that these construction deficiencies be mitigated to cope up with the housing demand.

The purpose of this study is to develop a supplementary construction policy guidelines and computational tool that will assist the users in mitigating these construction deficiencies. This study would identify the factors and assess its impact, identify the relative importance of each factor, and utilize ANN in developing the supplemental guidelines and computational tool.

## 2. Literature Review

There are different studies that identify the common causes of construction delays [10, 11, 12] and workmanship defects of various construction projects [13]. These are the studies that were utilized to identify the most common construction deficiencies that low-cost housing developments experience. There is also an increasing trend in the usage of Artificial Neural Network in construction management [11, 14]. There is usefulness to ANNs in carrying out a variety of prediction, classification, optimization, and modeling related tasks and it has significant benefits that make it a powerful tool for solving many problems in the field of construction management [15]. The utilization of ANN in the development of a computational tool will be helpful in validating the usage of ANNs specifically in housing projects.

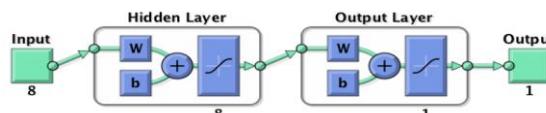
## 3. Methodology

### 3.1. Identification & Classification of the Factors

The sources of the factors are initially identified being the stakeholders involved in housing development projects. Different concepts of project management with relation to time and quality are reviewed and existing data from different related literatures that identifies the factors causing construction delays and workmanship defects are gathered to devise an assessment survey. The identified factors will comprise the assessment survey and will utilize a five-point Likert Scale. The survey was distributed to 60 professionals in the construction industry for impact assessment. Cronbach  $\alpha$  assessment is used validate the internal consistency of the questionnaires. Statistics of the gathered data is evaluated in preparation to the formulation of the ANN Model.

### 3.2. Artificial Neural Network Model and Relative Importance of Factors

The developed Artificial Neural Network model utilized a feed forward back propagation neural network model in predicting Construction Delays and Workmanship Defects rating. 4 ANN models were devised for this study in relation to the 4 data sets, Construction Delays Owner and Contractor Side, Workmanship Defects Owner and Contractor Side. Due to the similarity of the 4 data sets, identical criteria are employed to the ANN Model.



**Figure 1.** Design of the ANN Model

The training algorithm used for the model development of the 4 data sets was the Levenberg-Marquardt algorithm while the transfer function utilized is the hyperbolic tangent sigmoid function. The number of hidden neurons used for each model development is eight, equivalent to the number of input parameters such as the factors for each data set. The accuracy of the ANN models was evaluated based on the computed Pearson's Correlation Coefficient, Mean Squared Error and Mean Absolute Percentage Error. Utilizing the same ANN model, the study identified the relative importance of the factors using Garson's Algorithm.

### 3.3. Supplemental Guidelines and Computational Tool

Supplemental construction policy guidelines were devised based on the relative importance of each factor and the ANN model in preparation to the development of the computational tool. The computational tool was developed using the equations derived from the ANN model and will be distributed to professionals to validate its effectiveness in evaluating construction delays and workmanship defects impact and probability ratings on existing and future housing developments.

## 4. Results & Discussion

### 4.1. Identification & Classification of the Factors

The sources of these factors are identified. These stakeholders are categorized as follows: Project Owner & Contractor [7].

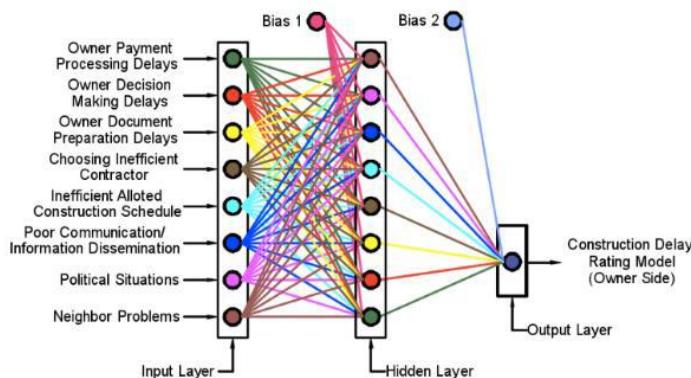
A study using artificial intelligence to predict risk delays in construction projects [12] identified 40 different factors that contribute to delays. Among the 40 factors, 8 are from the owner and 8 from contractors [12]. For this study, the researcher reclassified the delay sources according to the two major stakeholders involved in housing development projects. The Owner identified the following delay factors: A-Delay in payment processing, B-Owner decision making, C-Delay in preparing owner documents, D-the choice of inefficient contractor, E-the lack of a construction schedule, Factors F, G, and H are possible political situations, poor information dissemination, and neighbor

issues. Contractors: A-Problems with project planning and supervision, B-financial constraints C-poor internal communication, D-backlogs, E-poor labor, F-poor weather conditions, G-lack of construction materials, and H-unpredicted site conditions. This makes a total of 16 factors causing construction delays from owner and contractor side.

In a study on common workmanship defects in low-cost housing [16], 8 variables were identified as contributing to poor workmanship in project sites. The researcher used these sources to classify workmanship defects according to the two major stakeholders involved in housing development projects. The Owner identified the following workmanship defect factors: A-poor management supervision B-Refusal to pay for additional work, C-Approval of alternative materials, D-Shortening of work duration, E-Poor owner/developer maintenance, F-Uncared turned over units, G-Unsupervised modification of units, and H-poor maintenance. Contractors: A-Complicated subcontractor roles, B-Laborer lack of skill, C-Unsuitable construction equipment, D-Poor handling and storage of materials, E-Language/communication bar, F-Poor weather, G-vandalism, and H-nearby construction activities are all factors causing Workmanship Defects. This makes a total of 16 factors causing workmanship defects from owner and contractor side.

The 32 factors were assessed using a five-point Likert Scale. To assess the impact of each factor, 60 questionnaires were distributed to housing professionals with at least 5 years of experience. The Cronbach  $\alpha$  rating for construction delays (owner side) questionnaire was 0.896 indicating a good internal consistency and the Cronbach  $\alpha$  rating of the questionnaires for the construction delays (contractor side), work defects (owner and contractor side) were 0.935, 0.918, and 0.901 respectively, indicating an excellent internal consistency. The basic statistics for the 4 data sets are now established for the development of the ANN model.

#### 4.2. Artificial Neural Network Model



**Figure 2.** Architecture of Construction Delays Owner Side ANN Model

The developed Construction Delays rating model (Owner Side) has a  $R_{all}$  value of 0.99682 which is very high, and MSE and MAPE of 0.0040284 and 2.027%, respectively, which are very low, indicating that the model has high correlation and less error.

The correlation between the actual and the predicted Construction Delays rating model (Owner Side) yielded a Pearson's Correlation Coefficient value of 0.99675 which

suggests high correlation between the actual and the predicted Construction Delays rating model (Owner Side).

The derived prediction equation from the ANN model for the construction delay rating owner side using the weights and biases generated from the simulation will be utilized as the main driving components of the Computational Tool. Similarly, for the 3 remaining data sets Construction Delays Contractor Side, Workmanship Defects Owner & Contractor Side, same procedures were done to validate the reliability of the ANN model and derived equations.

#### *4.3. Relative Importance of Factors*

For the Construction Delay Owner side model, the ranking of importance of these factors is identified as Factor H < Factor E < Factor D < Factor F < Factor G < Factor C < Factor A < Factor B wherein Factor H – neighbor problems is the least influential factor with R.I. = 4.12895% and Factor B – owner decision making delays is the most influential factor with RI = 18.13171%.

For the Construction Delays Contractor Side model, the ranking of importance of these factors is identified as Factor F < Factor E < Factor D < Factor H < Factor A < Factor B < Factor G < Factor C wherein Factor F – poor weather conditions is the least influential factor with R.I. = 7.8220% and Factor C – poor internal communication is the most influential factor with RI = 17.2295%.

For the Workmanship Defects Owner Side model, the ranking of importance of these factors is identified as Factor F < Factor H < Factor D < Factor B < Factor E < Factor A < Factor G < Factor C wherein Factor F – tenants/ users lack of care for their units is the least influential factor with R.I. = 7.3241% and Factor C – owner's refusal to pay for additional works is the most influential factor with RI = 21.2203%.

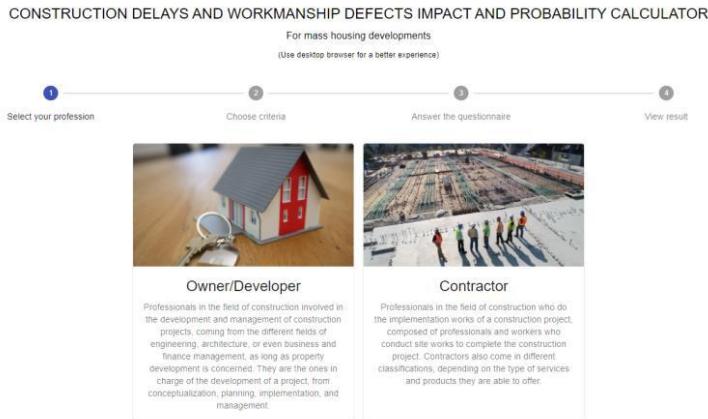
For the Workmanship Defects Contractor Side model, the ranking of importance of these factors is identified as Factor E < Factor H < Factor F < Factor A < Factor G < Factor B < Factor C < Factor D wherein Factor E – language/ communication barrier is the least influential factor with R.I. = 8.5079% and Factor D – poor handling and storage of materials is the most influential factor with RI = 23.7942%.

#### *4.4. Development of Supplemental Construction Policy Guidelines*

The supplemental guidelines can be accessed in the same web page as the computational tool. Each of the 32 factors were thoroughly defined for the user's reference. The guidelines will also define the required input from the user and the computational tool's output.

#### *4.5. Development of Computational Tool*

This study utilized Netlify, a web-based software development medium to develop the computational tool for this study. This web-based software utilizes Java code for the program of the tool.



**Figure 3.** Computational Tool Home Page

The computational tool will request the user to input their affiliation and select the construction deficiency he/she would like to assess, construction delay or workmanship defects. Then the computational tool will display the eight (8) factors associated to the choice of the user. The user will be requested to rate each of the factors according to its probability of occurring on his/her housing project using a five-point Likert scale. The computational tool will produce an output that will show the impact and probability rating that the user's construction project will encounter.

**Table 1.** Computational Tool Impact & Probability Rating

Output Range	Impact Rating	Probability Rating
Between 1 to 2	0-25% Impact	Unlikely to happen
Between 2 to 3	25-50% Impact	Less likely to happen
Between 3 to 4	50-75% Impact	More likely to happen
Between 4 to 5	75-100% Impact	Most definitely will happen

Impact rating derived from the Neural Network Model identifies the degree of impact of the factor the construction project will encounter. For the construction delays, it is the percentage increase from the original schedule, while for the workmanship defects, it is the weighted percentage of work items that will encounter defects. Probability Rating is derived from the user's input on the computational tool, directly related to the factors' relative importance. This will identify the probability of occurrence of construction delays or workmanship defects of a construction project given the users input.

To validate the computational tool's usability and effectiveness, the researcher introduced the computational tool to 10 experts in housing development projects. These professionals have an average years of work experience at 20 years. The researcher utilized a rubric assessment given to the professionals after they have tried using the computational tool. The computational tool was able to achieve an average rating of 4.6 out of 5 for the whole assessment. The computational tool's clarity and visibility had the highest rating of 4.9 while the user's requirements and the uniqueness had the lowest rating at 4.4. This assessment was helpful in identifying the usability of the tool and its possible room for improvements.

## 5. Conclusion

This study was able to utilize the implementation of artificial neural network in housing construction projects. The devised ANN model for the identified factors were able to derive the required equations for the computational tool. ANNs are modeled based on the input and output data and can be trained, having the potential in machine learning to always show updated results in utilizing new training examples. The developed supplemental construction policy guidelines and computational tool proved to be helpful in mitigating common construction deficiencies such as construction delays and workmanship defects having a 4.6 out of 5 rating in the rubric assessment given by the 10 professionals that tested the tool. The identified factors will give the users awareness regarding its impact on the project site, enabling them to address and prevent these deficiencies from occurring. However, this study is limited to the 32 identified factors that cause delays and work defects in low-cost housing projects, and the gathered data from the 60 respondents. The use of a web-based software for the development of the computational tool also allows it to be accessible to the users provided that they have access to a web browser and the internet. The computational tool also has the potential to be implemented in mobile-based app to further increase its accessibility.

## References

- [1] Build Build Build Philippine Infrastructure Transparency Program, [build.gov.ph](http://build.gov.ph)
- [2] Revised Implementing Rules & Regulations for BP220 – HLURB 2008
- [3] Housing Article, Philippine Board of Investments, <https://boi.gov.ph/ufaqs/housing/>
- [4] Mission and Vision – National Housing Authority, [nha.gov.ph](http://nha.gov.ph)
- [5] PSA Census of Population and Housing, <https://psa.gov.ph>
- [6] L. Olofsson and S. Truong (2007), Sustainable Housing in Navotas, Philippines – A Minor Field Study on Low-income Housing in Disaster Prone Areas
- [7] E. Adinyira, D. Ahadzie, and T. E. Kwofie (2013), Determining the Unique Features of Mass Housing Projects (MHPs)
- [8] G. J. Ritz and S. M. Levy (2013), Total Construction Project Management, Second Edition
- [9] B. Rajbhandari (2017), Construction Management in Mass Housing
- [10] N. Braimah (2013), Construction Delay Analysis Techniques – A Review of Application Issues and Improvement Needs
- [11] R. N. Macariola and D. L. Silva (2020), Coping with the Information Age: Development of a Data Flow Diagram-Based Knowledge Management System for Mitigating Delays for Construction
- [12] Z. M. Yaseen, Z. H. Ali, S. Q. Salih, and N. Al-Ansari (2020), Prediction of Risk Delay in Construction Projects Using a Hybrid Artificial Intelligence Model
- [13] H. Abdul Rahman, W. Chen, L. C. Wood, Y. M. Khoo (2014), Defects in Affordable Housing Projects in Klang Valley, Malaysia
- [14] K. M. de Jesus, D. L. Silva, B. S. Villaverde, E. M. Adina (2020), Hybrid Artificial Neural Network and Genetic Algorithm Model for Multi-Objective Strength Optimization of Concrete with Surkhi and Buntal Fiber
- [15] P. S. Kulkarni, S. N. Londhe, and M. C. Deo (2017), Artificial Neural Network for Construction Management: A Review
- [16] N. A. Othman and M. A. Othuman Mydin (2014), Poor Workmanship in Construction of Low-Cost Housing

# Research on the Impact of Customer Perceived Value of Haidilao Hot Pot on Consumer Behavior

Qiong LI<sup>1</sup>, Anlan LI

*Department of Economic Management, Wuhan Huaxia University of Technology,  
Wuhan, P.R.China*

**Abstract.** The paper takes Haidilao Hot Pot as the object through the collection of first-hand data, and analyzes the relationship between customer perceived value in functional value, emotional value, convenience value and social value on consumer behavior. Through factor analysis and variance analysis, it can be found that the four dimensions of customer perceived value have a significant positive impact on customer satisfaction, and customer satisfaction also has a significant positive impact on customer loyalty. Among them, functional value is the most important and obvious factor influencing customer purchasing decisions. The paper collects relevant data to complete the relevant content in the case of a global epidemic, hoping to provide certain guidance for the market-oriented adjustment of catering companies that are most affected by the epidemic.

**Keyword.** Catering; customer satisfaction; customer perceived value; consumer behavior; customer loyalty

## 1. Introduction

With the changes in the economic background of the times and the arrival of the experience economy era, what people buy is no longer the product itself but their expectations. Consumers hope to realize certain customer value in the transaction process, and the essence of customer value is customer perception, that is, customers' subjective perception of the interaction process and results of a certain enterprise.

Foreign scholar Zeithaml (1990) believes that when companies design, create, and provide value for customers, they should start from customer orientation, and take customer perception of value as the decisive factor[1]. Kolter (1969) and Levy proposed at the first time that "customer satisfaction depends on perceived value", which has attracted widespread attention abroad[2]. Then McDougall et al. (2000) further demonstrated the important role of customer value in business management in the research center of typical service industries, and concluded that there is a positive correlation between customer perceived value and customer satisfaction[3]. Li Man

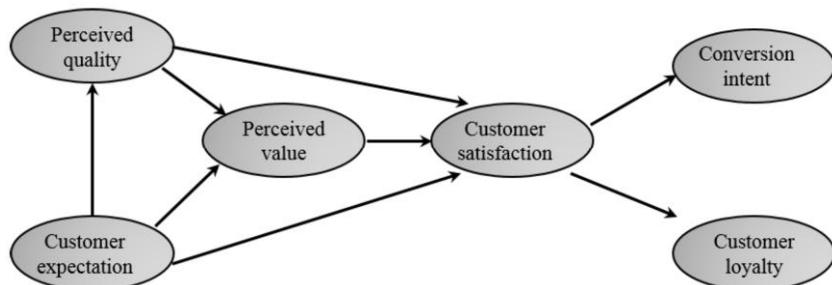
<sup>1</sup> Corresponding Author, Qiong LI, Department of Economic Management, Wuhan Huaxia University of Technology, Wuhan, P.R.China; E-mail: liqiong913@163.com; This paper was the research on the Influence Mechanism of the EWOM Information Type and Quality on the Consumer's purchase intention in the cultural consumption industry - Hubei film industry as an example (No. 18G143) by the General Humanities and Social Sciences Projects of the Ministry of Education 2018.

(2018) believes that quality, the role of brand, and the importance of customer value to the perception of customer experience are closely linked[4]. At the same time, related research shows that customer perceived value not only directly affects customer behavior, but also indirectly affects customer behavior through customer satisfaction. Zeithaml's research model points out that customer perceived value is an intermediate variable that affects quality and customer behavior, that is, customer Perceived value has a direct impact on customer behavior[5]. Domestic scholars Bai Changhong (2001) believe that customers always pursue maximum value, and then slowly modify their expected value from the accumulation of consumer experience, which indirectly affects customer satisfaction and behavioral tendencies[6]. The paper takes Haidilao, a well-known domestic catering company as the object, and encountered the global new crown epidemic during the research process. The collected data can better reflect the impact of customer perceived value on consumer behavior, and the customer perception of the catering industry in a specific market environment. The shaping of value has good theoretical and practical significance.

## 2 Theoretical models and indicators

### 2.1 Model framework

Claes For nell summarized the research results of customer satisfaction. CSI is referred to as the Customer Satisfaction Index for short, and the American Customer Satisfaction Index (ACIS) is also based on this[7].



**Figure 2.1** U.S. Customer Satisfaction Index Model Diagram(ACIS)

The perceived value of Haidilao studied in this article is that consumers recognize and choose the products of catering companies, and explain the perceived value of the value of catering products. Specifically, perceived value is the measurement of customers between perceived benefits and perceived costs[8]. Early scholars also believed that perceived value includes utilitarianism and hedonism. Since then, researchers have continuously deepened and expanded the research content of perceived value[9].

Therefore, this article proposes the following hypotheses: Hypothesis 1: Customer perceived value has a significant positive impact on customer satisfaction; Hypothesis 2: Customer satisfaction has a significant positive impact on customer loyalty; Hypothesis 3: Customer perceived value has a significant positive impact on customer loyalty.

## 2.2 Model variable definition and measurement indicators

### 2.2.1 Customer Perceived Value

The core factor of customer perceived value is the measurement between perceived benefit and perceived effort. The early definition of customer loyalty believes that only by excluding accidental factors, some deviation factors caused by convenient purchase and consumption, and the special case of customer multi-brand loyalty, continuous consumption purchase behavior can be used as a measure of customer loyalty. Customer satisfaction is that the actual expectations of customers are higher than the expectations of the customers themselves. This article makes corresponding evaluations of customer satisfaction from the following three aspects:

- (1)The customer's reaction to their product or service evaluation during the consumption process; (2)Customers' buying experience of products or services exceeds their expectations; (3)Customer's own consumption experience, which is consumption evaluation

### 2.2.2 measurement standard

**Table 2.1** Items for measuring customer perceived value

Measurement	Measurement item
customer perceived value	A01 Different feelings when entering a restaurant
	A02 A sense of intimacy from the restaurant staff
	A03 The pleasure of ordering and checkout speed
	A04 The comfort level of the store environment
	A05 Ease of parking
	A06 Location is close/multiple points
	A07 Reasonable price
	A08 Food safety
	A09 Complete food variety
	A10 Service from staff in time
	A11 Clean tableware
	A12 Delicious dishes
Functional value	A4; A11; A12; A13
Emotional value	A10; A2; A1
Convenience value	A3; A5; A6; A7
Social value	A8; A9

**Table 2.2** Measurement items of customer loyalty

Measurement	Measurement items
customer loyalty	B1 I often eat in this restaurant
	B2 Would recommend this restaurant to a friend
	B3 Compared with other restaurants, Haidilao's products and services are my first choice
	B4 Will continue to visit Haidilao

**Table 2.3** Measurement items of customer satisfaction

Measurement	Measurement items
customer satisfaction	C1 Haidilao products and services are excellent value for money
	C2 Haidilao's corporate image is good
	C3 Haidilao's services and products often exceed my imagination
	C4 Overall very satisfied with Haidilao

### 3. Questionnaire design and data analysis

#### 3.1 Data collection and collation

- 1) Selection of samples. The subject of this questionnaire is the customers of Haidilao, a branch of Haidilao Chain Enterprise in Wuhan City, Hubei Province.
- 2) This question is to study the impact of customer perceived value on consumer behavior and the impact of the CIS model based on customer perception. In the actual survey process, 200 questionnaires were issued and 179 valid questionnaires were collected. The recovery rate was 89.5%.

#### 3.2 Data analysis

##### 3.2.1 Questionnaire reliability analysis

It can be seen from Table 3.1 that the reliability coefficient value is 0.784, which is greater than 0.5, which indicates that the reliability of the research data is relatively high and suitable for questionnaire analysis.

##### 3.2.2 Questionnaire validity analysis

Table 3.2 shows the KMO sample measurement and Bartley sphere test results of customer perceived value.

##### 3.2.3 factor analysis

Through analyzing the five dimensions of customer perceived value, the reliability is high.

**Table 3.1** Trust level analysis

Name	Measured variable	$\alpha$ coefficient deleted	Cronbach's $\alpha$ coefficient
Customer Perceived Value	0.764	0.898	
Customer loyalty	0.743	0.842	0.784
Customer Satisfaction	0.634	0.753	
Standardized Cronbach's $\alpha$ coefficient: 0.784			

**Table 3.2** KMO and Bartlett test results

KMO sampling is an appropriate measure		.762
Bartlett's sphere test	Chi-square read above	457.641
	Degree of freedom	21
	Significance	.000

**Table 3.3** Customer Perceived Value Factor Analysis Results

	Component			
	Emotional value	Convenience value	Functional value	Social value
A01 Different feelings when entering a restaurant	.743	.204	.052	.324
A02 A sense of intimacy from the restaurant staff	.735	.279	.173	.192
A03 The pleasure of ordering and checkout speed	.114	.529	.413	.308
A04 The comfort level of the store environment	.058	.406	.688	.145
A05 Ease of parking	.247	.698	-.034	.008
A06 Location is close/multiple points	.153	.799	.157	-.045
A07 Reasonable price	.324	.634	.104	.208
A08 Food safety	.123	.040	.013	.584
A09 Complete food variety	.668	.072	.014	.865
A10 Service from staff in time	.253	.023	.014	.067
A11 Clean tableware	.447	.012	.083	-.106
A12 Delicious dishes	-.198	.084	.678	.420
Cronbach	.727	.702	.829	.835

**Table 3.4** Variance contribution

Factor naming	Functional perceived value	Convenience perception value	Sentimental value	Socially perceived value
Variance contribution	30.17%	11.023%	8.87%	8.67%
Cumulative variance contribution	58.733%			

### 3.2.4 regression analysis

(1) This section discusses the impact of several dimensions that affect perceived value on customer satisfaction from the perspective of customer perceived value. It uses customer satisfaction as the calibration standard and the predictive variable is the regression analysis of customer perceived value.

**Table 3.5** Overall model parameters

Model	R	R.Square	Adjuest R Square	Std.Error of the Estimate
1	.886(a)	.773	.783	.960

**Table 3.6** Regression coefficients and t test

Mode	Unstandardized Coefficients		Standardized Coefficients	t	Sig
	B	Std.Error	Beta		
I(Constant)	-.263	.645		-.408	.673
Functional value	.865	.042	.683	21.254	.000
Convenience value	.358	.032	.337	10.438	.001
Emotional value	.108	.032	.117	25.622	.004
Social value	.082	.052	.094	.10.511	.005

It can be seen from the statistical results of 3.5 that all dimensions of customer perceived value can be used as influencing factors of customer satisfaction.

The regression equation is:

$$CS=0.683F+0.117C+0.091E+0.337S$$

Through regression analysis, we know that all dimensions of perceived value have an impact on perceived satisfaction. Hypothesis 1 has been verified. Among them, functional value has the most obvious impact and emotional function is the weakest.

(2) This section discusses the impact of customer satisfaction on customer loyalty from the perspective of customer perceived value. It uses customer satisfaction as the school standard and the predictive variable for the regression analysis of customer loyalty value.

**Table 3.7** Overall model parameters

Model	R	R.Square	Adjuest R Square	Std.Error of the Estimate
1	.926(a)	.858	.856	.790

**Table 3.8** Regression coefficient and t test

Mode	Unstandardized Coefficients		Standardized Coefficients	t	Sig
	B	Std.Error	Beta		
I(Constant)	-.263	.645		3.252	.001
Customer Satisfaction	.924	.023	.925	42.074	.000

From 3.7 statistics, it is concluded that customer satisfaction can be used as an explanatory variable for loyalty. The regression equation is: CL=0.925CS

Through regression analysis, we know that customer satisfaction has a positive effect on customer loyalty. Hypothesis 2 is verified.

This section discusses the impact on customer loyalty from several dimensions of customer perception. The study uses perception as a variable and loyalty is the school standard for regression analysis.

**Table 3.9** Overall model parameters

Model	R	R.Square	Adjuest R Square	Std.Error of the Estimate
1	.897(a)	.803	.801	.930

**Table 3.10** Regression coefficient and t test

Mode	Unstandardized Coefficients		Standardized Coefficients	t	Sig
	B	Std.Error	Beta		
I(Constant)	-1.762	.623		-2.843	.005
Functional value	.812	.040	.643	20.254	.000
Convenience value	.298	.034	.280	9.438	.000
Emotional value	.125	.032	.118	3.922	.000
Social value	.147	.055	.079	2.611	.004

It can be seen from the statistical results of 3.9 that all dimensions of customer perceived value can be used as influencing factors of customer loyalty. Hypothesis 3 is verified.

The regression equation is:  $CL=0.643F+0.280C+0.118E+0.379S+i$  (i is a constant)

Through regression analysis, it is known that loyalty has an impact on the various dimensions of perceived value, of which functional value has the most obvious impact and social function is the weakest.

## 4. Conclusion and Outlook

### 4.1 Conclusion study

Based on survey interviews, this paper uses factor analysis to demonstrate the relationship between customer perceived value and customer satisfaction and customer loyalty. It is found that the four dimensions of customer perceived value have a significant impact on customer satisfaction, with functional value being the most significant. Emotional value is the weakest. It can be seen that to a certain extent, the perceived value of the function of the company determines the purchase decision and recommendation behavior of customers. Then, starting from several dimensions of customer perception, and using perception as the variable loyalty as the school standard, the regression analysis found that the various dimensions of customer perception value have an impact on customer loyalty. Among them, functional value has the most obvious influence, and social functional value is the weakest. Companies in a competitive industry try their best to satisfy customers first. If the company's products

and services are too general and do not make customers feel that they have obtained a higher consumption value, it will not be easy to attract customers to buy again.

#### 4.2 Management advice

##### 1) Service Packaging Strategy

Through research, we also know that perceived value is positively related to service satisfaction. The quality of the catering environment and the concept of good catering are all helpful to enhance the perceived value of consumers.

##### 2) Provide personalized services to meet the needs of different customers

In the era of perception economy, catering companies should give full play to their specialties and apply "personalized service marketing strategies" so as not only to meet the needs of different customers, but also to enhance the perceived value of customers.

##### 3) Improve customer satisfaction

Adjust product strategy in time, meet customer needs in a targeted manner, improve service quality, and increase satisfaction.

##### 4) Deliver to consumers with high-quality products and high-quality services

It is not only necessary to improve the quality of products to convey a good emotional impression to consumers, but also to give consumers the best quality and best services and products to become the leader of the enterprise.

##### 5) Propose strategies to maintain customer loyalty

Reasonably use the enterprise system to create consumer loyalty to the enterprise, establish consumer target groups, establish differentiated service marketing, and provide customers with high service quality.

## References

- [1] Zeithaml V.Parasuraman A, Berry. Delivering quality service, balancing customer perceptions and expectations[M]. New York: The Free Press,1990:30-32.
- [2] Kotler,P.and Levy,Broadening The Concept of Marketing into Nonprofit Organizations[J],Journal of Marketing,1969(33):10-15.
- [3] McDougall,S.,Curry,M.B.&Curry,M.Exploring the effects of icon characteristics on user performance:The role of concreteness,complexity and distinctiveness[J]. Journal of Experimental Psychology: Applied,2000,6,291-306.
- [4] Li Man, A Brief Analysis of the Relationship between Customer Perceived Value and Perceived Quality, Brand Image, and Customer Experience[J], Productivity Research, 2018, 11(2): 36-39
- [5] Zeithaml,Valarie A.Consumer Perceptions of Price, Quality, and Value:A Means-End Model and Synthesis of Evidence[J]. Journal of Marketing, 1998,(3):16-254
- [6] Bai Changhong, Western Customer Value Research and Its Practical Enlightenment[J], Nankai Management Review, 2001, (2): 72-74.
- [7] Chen Jun, Chen Xinlong, Sun Yimin, Xin Leyao, Yu Zhouyuan. Research on Haidilao Restaurant Service Marketing Strategy [J]. Journal of Shanghai University (Social Science Edition), 2021:3-5.
- [8] Zhong Kai, Zhang Chuanqing. Research on Consumer Perceived Value on Online Purchasing[D]. Wuhan: Zhongnan University of Economics and Law, 2017.
- [9] Zeithaml.Consumer Perceptions of Price Quality and Value: A Means. Model ald Synthesis of Evidence[J]. Journal of Marketing, 1988, 52(3): 2-22.

# Study on Construction of Hydrological Monitoring System of the Yangtze River and Application of New Technologies on Flood Monitoring

Junya MEI<sup>1</sup>, Bo ZHOU, Qiong WU

*Bureau of Hydrology, Changjiang Water Resources Commission, Wuhan 430010,  
China*

**Abstract.** The flood of the Yangtze River has the characteristics of high peak, large quantity and long duration. The Yangtze River Hydrology Bureau summarizes and combs the complete business process chain of flood hydrological monitoring, and gradually constructs the Yangtze River flood hydrological monitoring system. Including station network layout, early warning response, monitoring technology, information processing, results output and other dimensions. The hydrological monitoring system of the Yangtze River flood has been gradually constructed and has been successfully applied in many flood basins. Especially under the special situation of COVID-19 epidemic situation in 2020 and the severe flood situation in the Yangtze River Basin, the scientific and practical nature and practicability of the hydrological monitoring system of the Yangtze River flood are further verified. In view of the shortcomings existing in the existing monitoring system, this paper looks forward to the frontier technologies involved in flood monitoring, and has a certain reference function for flood hydrological emergency monitoring.

**Keywords:** The Yangtze River, the Yangtze River Economic Belt, flood, hydrological monitoring system, big data, Beidou satellite, GF satellite, Internet+

## 1. Introduction

The Yangtze River originates from the Geladandong Glacier on the Qinghai-Tibet Plateau and is the largest river in China. It flows through 11 provinces and cities, including Qinghai and Tibet, and finally into the East China Sea via Shanghai. The main stream is about 6,300 km long, and the basin area is about 1.8 million km<sup>2</sup>, accounting for 18.8% of China's land area. The Yangtze River system is composed of more than 7,000 tributaries, of which 8 tributaries have a catchment area of more than 80,000 km<sup>2</sup>, including the Yalong River, the Jialing River and the Wujiang River. At its middle-lower reaches, the main lakes are the Poyang Lake and the Dongting Lake [1].

The Yangtze River, which runs through the east and west of China, is known as the "Golden Waterway." In 2016, China established a new development pattern of "one axis, two wings, three poles and multiple points" of the Yangtze River Economic Belt. This Belt covers 21% of the land area, and its population and total economic output both

---

<sup>1</sup>Corresponding author: Junya Mei, Bureau of Hydrology, Changjiang Water Resources Commission, Wuhan 430010, China; E-mail: 2280035532@qq.com.

exceed 40% of the whole country. Therefore, it has an important ecological position, strong comprehensive strength and huge development potential [2]. The “Yangtze River protection” is an important measure based on the strategy of the Yangtze River Economic Belt. It is very important for the development and construction of the Yangtze River Economic Belt to control floods of the Yangtze River.

The Yangtze River Basin is basically located in the East Asian subtropical monsoon area, most of which belong to the subtropical zone. There are many rainstorms in the Yangtze River Basin, and the rainstorm is widely distributed, and its flood is mainly caused by rainstorm. According to historical records, between 206 BC and 1911, 214 floods occurred in the Yangtze River, with an average of 10 years. In the middle of the 19th century, two major floods occurred in 1860 and 1870 in succession. There was a major flood about six years after 1911. The floods of 1931 and 1935 killed more than 140000 people. In 1954, the flood area flooded 3.17 million hm<sup>2</sup>, of cultivated land and killed more than 30, 000 people as a result of the disaster, and the Beijing-Guangzhou Railway was shut down for 100 days. Another flood occurred in 1998, in the upper and middle reaches of the Yangtze River in 2010, in the middle and lower reaches of the Yangtze River in 2016 and 2017, in the upper reaches of the Yangtze River in 2018 [3], in the middle and lower reaches of the Yangtze River in 2019, and in the Yangtze River in 2020. The flood of the Yangtze River has obvious regional characteristics, temporal and spatial characteristics and flood characteristics of high peak, large quantity and long duration. If we do not deal with it scientifically, it will cause serious flood and waterlogging disasters and seriously damage the safety of the lives and property of the people in the basin.

In order to successfully cope with the flood and ensure the sustainable development of the Yangtze River economic belt, it is necessary to control the flood reasonably through fine forecasting and dispatching. Taking the 2020 Yangtze River flood as an example, in mid-August, the main tributaries of Mintuo River and Jialing River in the upper reaches of the Yangtze River suffered seriously. The three Gorges Reservoir had the biggest flood since it became a reservoir, and the flood peak reached 75000m<sup>3</sup>/s at 8:00 on the 20th. Through fine forecast, combined operation of reservoir group in the upper reaches of the Yangtze River, cumulative flood storage of 19 billion m<sup>3</sup>, and reduction of flood peak (flood volume) once in 90 years (130 years) to 20 years (40 years), the flood control pressure of the basin has been significantly reduced, the opening of Jingjiang flood diversion area has been avoided, and great benefits of flood control and disaster reduction have been brought into play [4]. In addition to hydrological prediction and water engineering dispatching technology, the high precision technology behind these successful cases can not only be separated from scientific, efficient and systematic hydrological monitoring technology. Hydrological monitoring system provides efficient and accurate data sources for forecasting and dispatching system, and through real-time feedback between front and rear, constantly optimize survey layout and result analysis and processing, and cooperate with each other to provide support for flood and drought disaster prevention. Through the monitoring practice of the Yangtze River flood, the hydrological monitoring system of the Yangtze River flood has been gradually formed.

## 2. Construction of Hydrological Monitoring System of the Yangtze River on Flood Monitoring

### 2.1. Overall Architecture

The hydrological monitoring system of the Yangtze River includes the station network layout, early warning and response, monitoring technology, information processing and

result output, as shown in Figure 1.

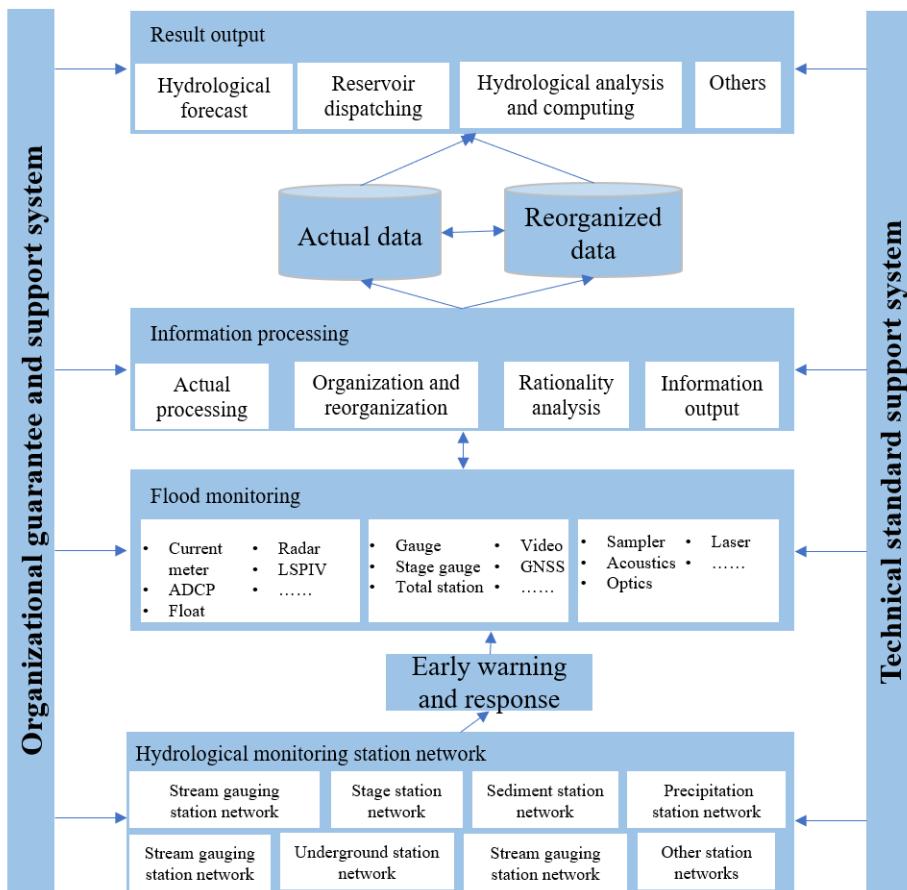


Figure 1. Structure diagram of flood monitoring system of the Yangtze River.

## 2.2. System Components

### (1) Hydrological monitoring station network

The hydrological station network is the “workplace” of hydrologists. After four large-scale station network planning and adjustment in 1956, the Yangtze River Basin has formed the current station network system with reasonable structure and scientific layout [5]. The Yangtze River Hydrology has set up more than 300 hydrological monitoring stations at the important water system nodes such as the main stream of the Yangtze River and the main tributaries, and uses "clairvoyance" to control the water and rain of the Yangtze River in real time. At the same time, with the hydrological and meteorological departments of the provinces and cities in the basin, reservoir management departments, information sharing, the Yangtze River flood control forecasting and dispatching system has gathered 30,000 water and rain stations in the basin.

## (2) Early warning response

In addition to the requirements of Hydrologic Survey Assignment, a special early warning and response mechanism was set up as the driving factor to carry out flood monitoring of the Yangtze River. When the driving conditions of Regulations on Flood Numbering of Major Rivers in China, Regulations on Flood Numbering of Major Trans-Provincial Tributaries of the Main Stream of the Yangtze River from Shigu to Cuntan Section and the Basin and Emergency Plan for Flood and Drought Disaster Prevention in the Yangtze River Basin are satisfied, flood monitoring of the Yangtze River shall be carried out immediately.

## (3) Monitoring technology

After many years of development, the Yangtze River Hydrology has formed a monitoring system of "combination of resident and patrol, priority of inspection and survey, automatic monitoring and reporting, emergency supplement" [6]. In the process of flood monitoring in the Yangtze River, in addition to the use of conventional monitoring technology, in recent years, it has gradually changed to non-contact and rapid monitoring means, and adopted the innovative model of multi-source data fusion to carry out monitoring and data rationality analysis and processing, so as to ensure the timeliness and accuracy of flood monitoring data. At present, according to the design concept of "one station, one policy", the Hydrological Emergency Forecast Plan of the Yangtze River Flood has been formulated to ensure the collection and submission of key elements such as water level and discharge in the event of severe floods. Recently, Beidou navigation satellite system, which has been successfully connected around the world, has also played an important role in flood control and disaster relief. As one of the core functions of Beidou, the short message function provides a guarantee for the transmission of key hydrological information in the harsh natural and network environment caused by heavy rain and flood in flood season.

## (4) Information processing

The hydrological industry is a typical data production industry. Under the Bureau of Hydrology, nearly 80% of the hydrological stations have accumulated more than 30 years of continuous hydrological data and a large quantity of flood monitoring data. The big data analysis method is used to mine the value of historical hydrological monitoring data, and provide support for the arrangement of major flood times of measurement, the alignment of water stage-discharge relationship, and the corresponding discharge reporting. Relying on the model of "Internet + hydrological monitoring", tasks management, instrument calibration, hydrological survey, data compilation and other links are effectively connected in the form of information flow, providing support for the information processing of Yangtze River flood monitoring [7-8]. Based on the Yangtze River hydrological monitoring and management platform and the hydrological data online compilation system, information processing is completed, data is collected into the real-time database and compilation database, and standard API interfaces are made available for the Yangtze River flood control forecasting and dispatching system and other platforms.

## (5) Result output

The results of flood monitoring are stored in the compilation database and real-time database, and provide basic data for hydrological forecast and hydrological analysis and calculation.

## (6) Support system

The successful implementation, the organizational guarantee and technical standard system of Yangtze River flood monitoring play a very important role as a foundation

support. The organizational guarantee covers work guarantee, traffic guarantee, communication guarantee, technical guarantee, emergency support guarantee, safety guarantee and training and drill guarantee, etc. The Technical Guidelines for Hydrologic Emergency Monitoring and other national and industrial technical standards guarantee the accuracy, rationality and suitability of the monitoring methods for large floods, and are an important means to ensure the quality of monitoring results.

### **3. Practice of the Yangtze River Flood Monitoring in 2020**

The Bureau of Hydrology carried out hydrological monitoring over the Yangtze River flood under the support of the hydrological monitoring system of the Yangtze River.

#### *3.1. Monitoring Preparation*

The hydrological stations of the Yangtze River Hydrology Bureau have a wide range of lines. In the face of the 2020 epidemic situation of COVID-19 and the severe flood control situation, the Yangtze River Hydrology Bureau has always adhered to both the epidemic prevention and control and flood control work, adopted extraordinary methods to make preparations before flood; innovated and adopted the working mode of "self-inspection and field spot inspection", developed a three-level network self-inspection system to complete flood inspection and rectification. Timely compilation and release of the Yangtze River extraordinary flood hydrological emergency forecast plan, the whole river organized more than 20 actual flood control forecasting exercises, successfully completed the hydrologic bureau level super standard flood forecasting exercises, for the smooth implementation of the 2020 Yangtze River flood has made technical preparations.

#### *3.2. Overview*

In 2020, under the influence of concentrated heavy rainfall, a flood occurred in the Yangtze River. Since June, eight major floods have occurred in Wujiang River, Jialing River and the upper reaches of the Yangtze River. The gaochang station in Minjiang River has the biggest flood since the station was built in 1939; the Fushun station in Tuojiang River has the biggest flood since the station was built in 2001, which is the biggest flood since 1981; the Beibei station in Jialing River has the second largest flood since it moved to the station in 2007; and the second largest flood has occurred since the station was built in 1939 in the main stream of the Yangtze River. The largest flood has occurred since the completion of the three Gorges Reservoir, and the flood peak reached  $75000m^3/s$  at 8: 00 on August 20, as shown in Table 1.

**Table 1.** Basic data about the numbered floods of the Yangtze River and main tributaries in 2020.

No.	Time	Flood situation	Cause
The Wujiang River 1	6-22 15:00	The water stage at the Wulong Station rose to 193.08m	Affected by the intensified heavy rainfall, the incoming water increased rapidly
The Jialing River 1	8-12 12:00	Above the warning water stage at the Xiaohiba Station	Affected by heavy rainfall and incoming water from the upper reaches, incoming water from the Fujiang, Tuojiang and Minjiang Rivers at the upper reaches of the Yangtze River increased rapidly
The Jialing River 2	8-16 13:00	Above the warning water stage at the Xiaohiba Station	Affected by heavy rainfall, incoming water from the Fujiang River a tributary of the Jialing River at the upper reaches of the Yangtze River increased rapidly
The Yangtze River 1	7-2 10:00	Incoming discharge to the Three Gorges Reservoir 50,000m <sup>3</sup> /s	Affected by heavy rainfall, incoming water from the main stream of the Yangtze River, the Fujiang River and the Three Gorges area increased rapidly
The Yangtze River 2	7-17 10:00	Incoming discharge to the Three Gorges Reservoir 50,000m <sup>3</sup> /s	Affected by heavy rainfall, incoming water from the main stream of the Yangtze River and the Three Gorges area increased rapidly
The Yangtze River 3	7-26 14:00	Incoming discharge to the Three Gorges Reservoir 50,000m <sup>3</sup> /s	Affected by heavy rainfall, incoming water from the main stream of the Yangtze River, the Jialing River, the section between the Xiangjiaba and Cuntan Stations, and the Three Gorges area increased rapidly
The Yangtze River 4	8-14 5:00	Discharge at the Cuntan Station increased to 50,900 m <sup>3</sup> /s	Under the influence of heavy rainfall, floods above the guaranteed stage occurred in Tuojiang River and Fujiang River, and floods above the warning stage happened in the Minjiang River and the section between Luzhou and Cuntan
The Yangtze River 5	8-17 14:00	Discharge at the Cuntan Station increased to 50,400 m <sup>3</sup> /s	Under the influence of heavy rainfall, floods above the warning stage occurred in the tributaries the Minjiang River, the Tuojiang River and The Jialing River, and floods above the guaranteed stage happened in the Fujiang River

### 3.3. Flood Monitoring

Based on the flood hydrological monitoring system, the Bureau of Hydrology successfully dealt with the eight numbered floods in 2020, and completed the change process of hydrological elements such as water stage, discharge and sediment, which provided a solid support for hydrological forecast and hydrological analysis and calculation. Taking the discharge monitoring as an example, during the eight numbered floods, the 12 controlling water system nodes such as Cuntan, Yichang, Hankou, Jiujiang, Datong, Xiaoheba and Gaochang adopted Acoustic Doppler Current Profiler (ADCP), current meter method, floats method, electric flow meter method, and other methods to complement each other; from June to August, the discharge was measured in

a total of 651 times (see Table 2). The times of measurement of ADCP, current meter method, floats method, and electric flow meter method accounted for 57%, 38%, 4% and 1% respectively. It can be seen that rapid discharge monitoring technologies such as ADCP became the main technical means of flood monitoring.

The discharge times of measurement of the stations related to the Yangtze River flood monitoring from June to August are shown in Figure 2. It can be seen from the figure that the total times of measurement of Cuntan Station are up to 78; According to the monthly distribution, there are 139 times in June, 240 times in July and 272 times in August, with the largest number in August, mainly due to the occurrence of 8 numbered floods in August, including 2 numbered floods in Jialing River and 2 numbered floods in Yangtze River. Taking Cuntan in the main stream of the Yangtze River as an example, in the five numbered floods, scientific times of measurement were arranged according to the station characteristics and flood change characteristics of Cuntan Hydrological Station, so as to completely control the discharge change process (see Figure 3).

Rational analysis shows that the water stage and discharge process of each station are corresponding, the process comparison of upstream and downstream stations is reasonable, the flood peak water stage and peak discharge time are corresponding, the water volumes of upstream and downstream are balanced, and the flood survey results are reliable. See Table 3 for the statistics of flood characteristic values of all relevant stations.

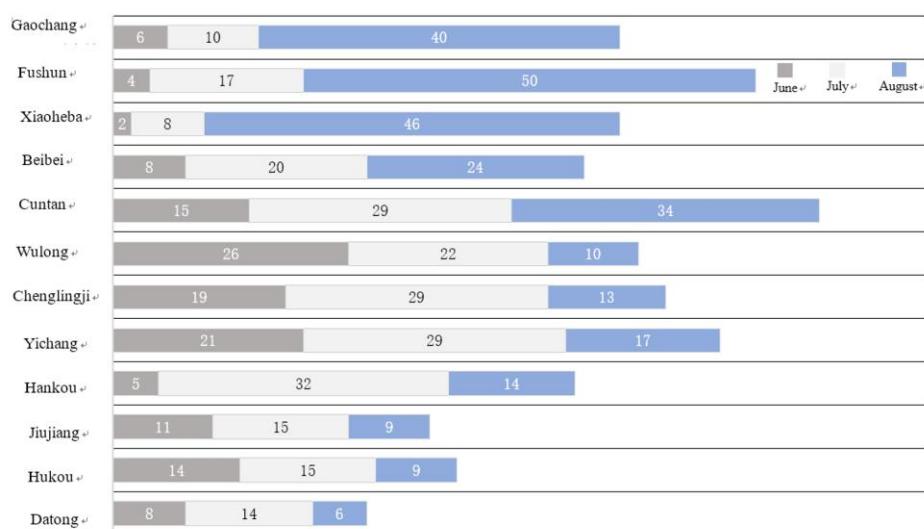
**Table 2.** Statistical data of discharge times of measurement at stations related to the Yangtze River flood from June to August in 2020.

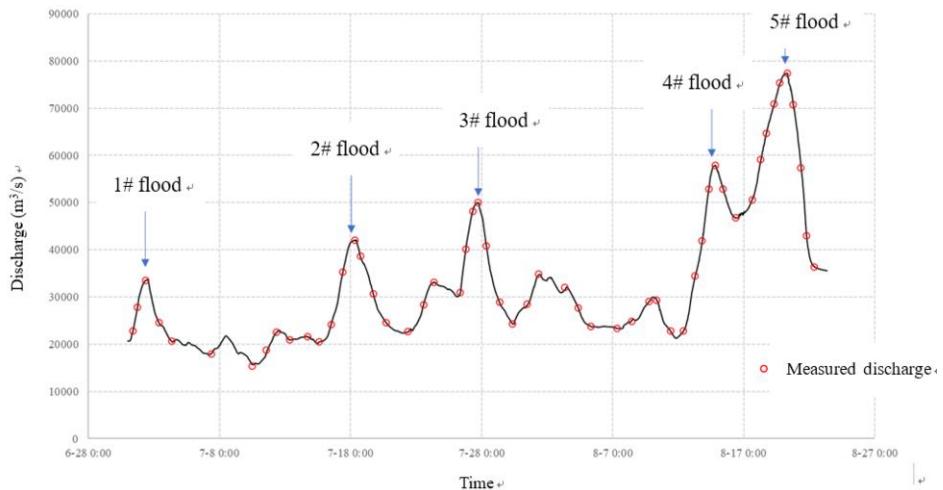
Observation station	River	Observation method			
		ADCP	Current meter	Float	Electric flow meter
Gaochang	The Minjiang River		55	1	
Fushun	The Tuojiang River		71		
Xiaoheba	The Fujiang River	25		23	8
Beibei	The Jialing River		52		
Cuntan	The Yangtze River	75	3		
Wulong	The Wujiang River		58		
Chenglingji	The Yangtze River	61			
Yichang	Dongting Lake Waterway	56	11		
Hankou	The Yangtze River	51			
Jiujiang	The Yangtze River	35			
Hukou	Poyang Lake Waterway	38			
Datong	The Yangtze River	28			

**Table 3.** statistical data of characteristic values from the related stations during the Yangtze River flood in 2020.

Observation station	River	Discharge		Stage		Characteristic stage of flood control		Flood situation	
		Measu red max.	Time	Corresponding stage	Measu red max.	Time	Waring stage	Guaranteed stage	Above warning stage
Gaochang	Minjiang	37500	8-18	291.07	291.11	8-18	285.00	288.00	6.11
Fushun	Tuojiang	10400	8-18	274.08	274.11	8-18	272.30	274.00	1.81
Xiaoheba	Fujiang	23900	8-17	243.36	245.35	8-17	238.00	240.00	7.35
Beibei	Jialing	32800	8-19	200.00	200.23	8-19	194.50	199.00	5.73
Cuntan	Yangtze	77400	8-20	191.56	191.62	8-20	180.50	183.50	11.12
Wulong	Wujiang	15000	6-22	196.27	196.28	6-22	193.00	199.50	3.28
Yichang	Yangtze	51500	8-21	53.45	53.51	8-21	53.00	55.73	0.51
Chenglingji	Dongting Lake	32500	7-12	34.56	34.74	7-28	32.50	34.55	2.24
Hankou	Yangtze	61900	7-29	28.44	28.77	7-12	27.30	29.73	1.47
Jiujiang	Yangtze	66000	7-29	21.72	22.81	7-12	20.00	23.25	2.81
Hukou	Poyang Lake	24000	7-11	22.18	22.49	7-12	19.50	22.50	2.99
Datong	Yangtze	83400	7-12	16.08	16.24	7-13	14.40	17.10	1.84

Note: 1. Water stage (in m), discharge (in  $m^3/s$ ); 2. The water stage datum of each station is its stationary datum.

**Figure 2.** Diagram of discharge times of measurement distribution at stations related to the Yangtze River flood from June to August in 2020.



**Figure 3.** Diagram of times of measurement arrangement of Cuntan station discharge process control during 5 Yangtze River numbered flood in 2020.

#### 4. Conclusion

In view of the characteristics of high peak, large volume and long duration of the Yangtze River flood, safeguarding the Yangtze River and controlling its floods is very important to develop and construct the Yangtze River Economic Belt. In order to cope with deluges successfully and guarantee the sustainable development of the Yangtze River Economic Belt, it is necessary to control them rationally through precise forecast and water dispatching. By continuous summarization and practice of monitoring over the previous floods in the Yangtze River, the Bureau of Hydrology has gradually established a scientific, efficient and complete Yangtze River flood hydrology monitoring system. This system can provide efficient and accurate data sources for the forecasting and dispatching systems to complement each other to provide support for prevention of floods and droughts. In the monitoring of many river basin floods, especially in the 8 numbered floods in 2020, it was successfully applied to collect continuous, reliable, complete water and sediment data. However, in the monitoring practice, some obvious weaknesses were also reflected, such as insufficient monitoring capacity, insufficient intelligence, and insufficient integration mechanism of various monitoring results. In order to further improve the hydrological monitoring system of the Yangtze River flood and enhance its monitoring capability and intelligence level, relevant research can be carried out in the following aspects:

(1) Study on the application of satellite remote sensing in flood monitoring. With the continuous development of satellite remote sensing technology, its time and space resolutions have been greatly increased. Satellite remote sensing data, combined with data of water level, flow rate and flow section from hydrological stations, can be used to carry out inversion model research based on supervised learning, such as satellite water level measurement and discharge computation, as well as quantitative accuracy

evaluation, as an emergency supplement in case of severe floods where normal hydrological survey cannot be conducted by conventional means.

(2) Study on the application of machine learning in flood monitoring. The Bureau of Hydrology has collected a great deal of flood survey data. Therefore, the survey scheme can be continuously optimized based on machine learning algorithms such as Logistic Regression, Support Vector Machine and Random Forest, to research the most optimal strategy in survey method selection, vertical layout, measurement point selection and duration selection during the period of high flood, and continuously improve the intelligence level of flood monitoring.

## Acknowledgments

This work is supported by National Key R & D Program of China (2018YFC1508002).

## References

- [1] Encyclopedia of Rivers and Lakes in China Compilation Committee. Encyclopedia of Rivers and Lakes in China· The Yangtze River Volume (1st) [M]. Jan. 2010 1st version. China Water & Power Press, 2010.
- [2] Deng S, Mei JY, Zhou B et al. Research on hydrological emergency monitoring scheme and technology of Barrier Lake. IOP Conf. Ser.: Earth Environ. Sci. 2019, 344(2019):012102.
- [3] Chen M. Construction and achievements of flood control, draught relief and disaster reduction systems in the Yangtze River Basin China. Flood & Drought Management. 2019(10), 29:36-42.
- [4] Cheng HY. Flood forecasting system its role in project networks China. Water Resources. 2020, 17: 11-3.
- [5] He H. China gauging station network Advances in Water Science 2010, 21: 460-5.
- [6] Wang J et al Modern Hydrologic Monitoring Technology [M]. Dec. 2016 1st version. China Water & Power Press, 2016.
- [7] Zhou B et al. Research and practice of "internet + hydrological monitoring" of the Yangtze River. IOP Conf. Ser.: Earth Environ. Sci. 2019, 344: 012104.
- [8] Wang J. Research on the hydrological monitoring system of the Yangtze River based on "Internet+ [J]. Technology and Economy of Changjiang. 2018, 2:70-4.

# Management Competency Model: Predictive Neural Network Approach for Empirical Components of Construction Project Proficiency

Dante L. SILVA<sup>a</sup><sup>1</sup>, Kevin Lawrence M. De JESUS<sup>b</sup>, Bernard S. VILLAVERDE<sup>a</sup>,  
Edgar M. ADINA<sup>a</sup>

<sup>a</sup>*School of Civil, Environmental and Geological Engineering, Mapua University, 658  
Muralla St., Intramuros, Manila 1002, Philippines*

<sup>b</sup>*School of Graduate Studies, Mapua University, 658 Muralla St., Intramuros, Manila  
1002, Philippines*

**Abstract.** Management competencies are skills that incorporate the understanding, proficiencies and qualities essential for successful performance. Individuals in the top of the organizational hierarchy presents himself being an effective leader by immersing to a readily difficult activity in the project. Molding the important management competencies was found to be hard since the efficacy of a competent construction manager is dependent on countless administrative aspects. The current study intended to offer a construction management competency for human resource development in construction companies. Utilizing this competency model, the company could increase its performance capacity and productivity. This study developed a competency theory and a hybrid predictive model with specific foci on construction managers. The Management Competency Framework Assessment Instrument was developed, following an approach through factor-item analytic mode. Furthermore, this research constructs a predictive performance model using Artificial Neural Network through the factors associated to successful management performance. A sensitivity analysis was also implemented to assess the relative importance of individual factors to the effective construction performance.

**Keywords:** Management competency model, artificial neural network, construction project proficiency, sensitivity analysis

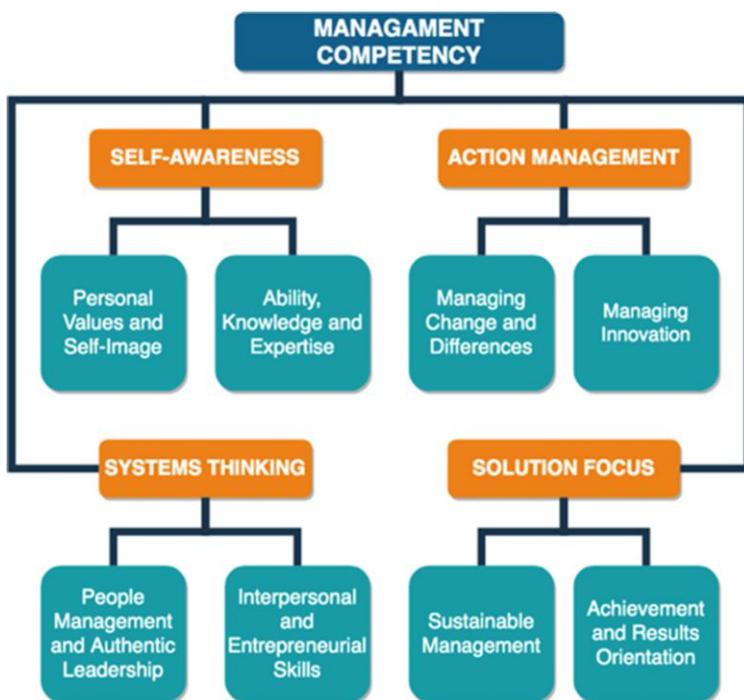
## 1. Introduction

One of the most difficult undertakings in a construction project is the selection of the most suitable individual for the job. Currently, project managers in the construction industry requires versatility and adaptability to multiple functions and roles wherein a key element is their knowledge and skills gained from years of experience and training [1,2]. The management competency is classified into four areas: self – awareness,

---

<sup>1</sup>Corresponding author: Dante L. Silva, School of Civil, Environmental and Geological Engineering, Mapua University, 658 Muralla St., Intramuros, Manila 1002, Philippines; E-mail: dantesilva2000@yahoo.com.

systems thinking, action management and solution focus as shown in Figure 1. Self-awareness is a management competency which refers to how a person analyzes situations carefully and makes rational judgments and logical decisions. Personal values and self – image (A1) and ability, knowledge and expertise (A2) were the subsets under this management competency area. Systems thinking is a driver of how thinking about problems and strategies should be approached within the organization. Under this management competency area, it includes people management and authentic leadership (B3) and interpersonal and entrepreneurial skills (B4). Action Management refers to the act, manner or practice of managing changes and innovations and handling differences. It includes managing change and differences (C5) as well as innovation management (C6). The last area of professional management competency is the solution focus which is the knowing the important results and focusing resources to achieve clear outputs. It involves sustainable management (D7) and achievement and results orientation (D8) [3].



**Figure 1.** Management competency areas.

Assessment of the competency for successful project management performance is a major challenge in the construction industry. Lack of project management competency may result to issues and problems in the costs, delays, labor productivity and safety in the construction site. The use of machine learning techniques such as Artificial Neural Network (ANN) may assist in addressing the issue of competency assessment [4].

The use of the machine learning techniques was steadily increasing over the years. In the construction management, the use of machine learning techniques such as neural network was implemented in prediction of safe work behavior [5], construction labor

productivity [6], construction project manager's motivators and human values [7] and change orders [8]. The use of neural network as prediction is due to its ability to model complex and non-linear functions and minimum statistical requirements.

This study established a neural network model for predicting construction project competency using different management competency subsets. Moreover, the relative importance of these subsets to the construction project management competency rating was calculated.

## 2. Methodology

### 2.1. Research Framework

The flow of the implementation of the study for the prediction of the management competency rating as well as calculation of the most influencing factors were displayed in Figure 2.

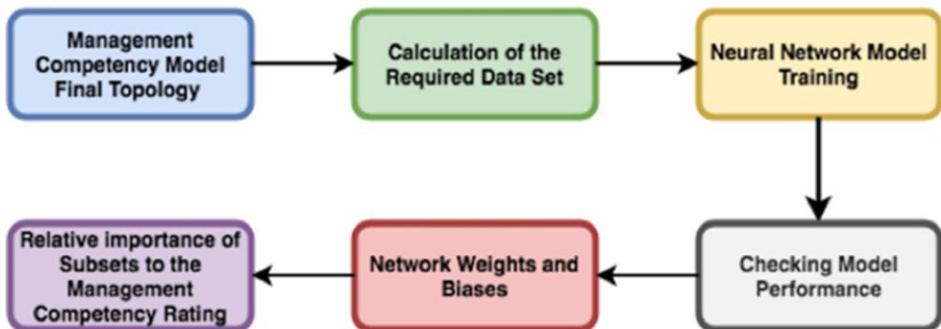


Figure 2. Framework of the study.

### 2.2. Data Collection

The number of respondents used in this study is established on the suggested topology of the ANN model wherein the quantity of hidden neurons is signified by “m” quantity of input parameters [9]. These suggested quantity of hidden neurons for an ANN model was used to calculate the range of the required number of data sets [10] as shown in equation (1) and (2).

$$MIN = HI + OH \quad (1)$$

$$MAX = H(I+1) + O(H+1) \quad (2)$$

where “I” is the quantity of input parameters, “H” is the quantity of hidden neurons and “O” is the quantity of output parameters. Using this equation, this study arrived with the 81 respondents required for the training of the network.

### 2.3. Neural Network Modelling

The management competency rating model was created using the MATLAB R2019a Neural Network Toolbox™. The data gathered from the respondents were partitioned into three groups: 70% were used for training, 15% for validation and 15% for testing [11]. Using the governing design of the model which is 8-8-1 (input neurons-hidden neurons-output neuron).

The datasets were trained and the developed management competency rating model was determined. The governing characteristics of the model is shown in Table 1.

**Table 1.** Governing characteristics of the management competency model.

Internal Characteristics	Value
Training Algorithm	Levenberg – Marquardt [12]
Transfer Function	Hyperbolic Tangent Sigmoid [12]
Number of Hidden Neurons	8
Number of Hidden Layers	2 (1 Hidden Layer and 1 Output Layer)
Performance Indicators	R, MSE and MAPE [13]

The developed model was obtained using different performance indicators as shown in equation (3) – Pearson’s Correlation Coefficient, equation (4) – Mean Squared Error (MSE) and equation (5) – Mean Absolute Percentage Error (MAPE).

$$R = \frac{\sum_{j=1}^n (x_j - \bar{x}_j)(z_j - \bar{z}_j)}{\sqrt{\sum_{j=1}^n (x_j - \bar{x}_j)^2 \sum_{j=1}^n (z_j - \bar{z}_j)^2}} \quad (3)$$

$$MSE = \frac{1}{n} \sum_{j=1}^n (x_j - z_j)^2 \quad (4)$$

$$MAPE = \frac{1}{n} \sum_{j=1}^n \frac{|y_j - \bar{y}_j|}{y_j} \times 100 \quad (5)$$

where the  $x_j$ ,  $y_j$  and  $z_j$  represents the actual value, determined value for “j” variable and predicted value, respectively while  $\bar{x}_j$ ,  $\bar{y}_j$  and  $\bar{z}_j$  indicates the average of the targeted real value, predicted output and predicted value by the ANN for parameter “j”, respectively [14].

### 2.4. Sensitivity Analysis

A sensitivity analysis was implemented to determine the influence of individual input parameter to the management competency rating and its variability. Since this study includes multiple input parameters, the sensitivity analysis is the principal tool for evaluation of the variation of the output and provided an explanatory analysis on the input parameter influence to the model output [15].

The Garson's Algorithm was utilized to assess the impact of input parameters individually to the management competency rating. The relative importance of input variables to the target variable is established on the weights of the connections which is given by Equation (6).

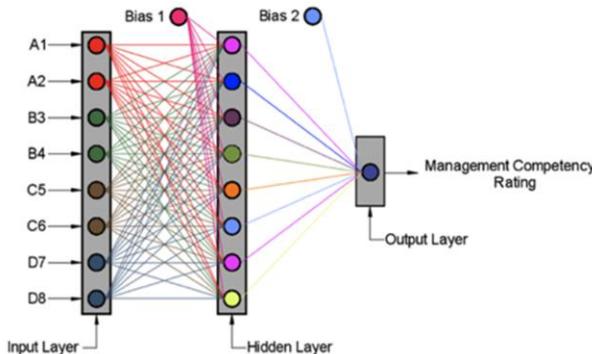
$$RI = \frac{\sum_{b=1}^n \left[ \frac{|w_{ab}|}{\sum_{i=1}^m |w_{ib}|} |w_{bc}| \right]}{\sum_{a=1}^m \left\{ \sum_{b=1}^n \left[ \frac{|w_{ab}|}{\sum_{i=1}^m |w_{ib}|} |w_{bc}| \right] \right\}} \quad (6)$$

where RI is the relative importance of the input parameters to the management competency rating, "a" is the quantity of input variables, "b" is the quantity of hidden neurons and "c" is the quantity of output variables [16].

### 3. Results and discussion

#### 3.1. Neural Network Modelling

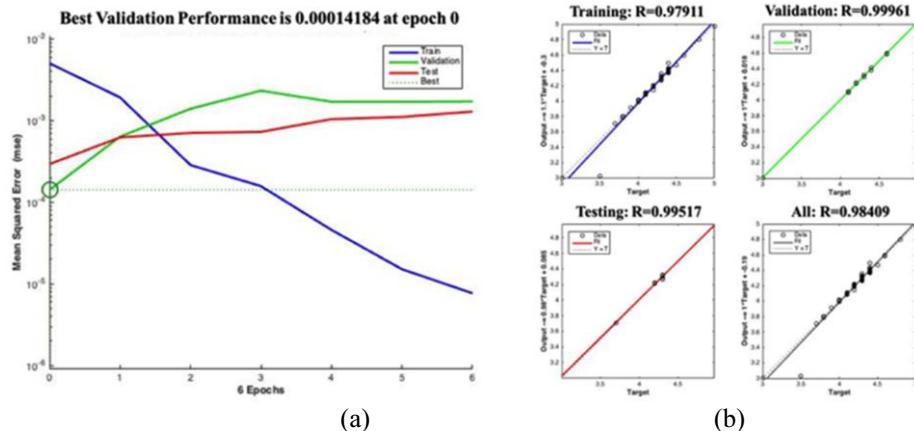
Neural network model development which is best suitable to for this study, is mainly relying on the architecture of the network. Using the related literature for neural network model development, the final topology and architecture of the model was identified and determined to be 8-8-1 (input-hidden-output) and shown in Figure 3.



**Figure 3.** Architecture of the management competency rating model.

The performance of the developed management competency model has an extremely high  $R_{all}$  value and very low MSE and MAPE. The performance and

correlation plots of the developed management competency model were shown in Figure 4a and 4b.

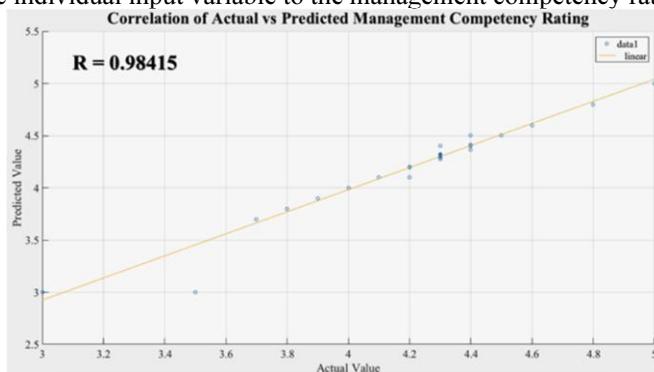


**Figure 4.** (a) Performance plot and (b) Correlation plots of the developed management competency model.

The performance of the developed management competency model comprises of plots from three distinctive phases of the model development – training, validation and testing phases which was pre-assigned in the Neural Network Toolbox as 70%, 15% and 15%, respectively, from the entire dataset. The regression values for the developed management competency model were 0.97911, 0.99961, 0.99517 and 0.98409, for training, validation and testing, respectively. The performance of the model was MSE equal to 0.00014184, MAPE of 0.66393559% and  $R = 0.98415$  between actual and predicted management competency rating as shown in Figure 5.

### 3.2. Sensitivity Analysis

In determining the influence of individual input variable to the management competency rating, sensitivity analysis using Garson's Algorithm was executed. Table 2 presents the product weights of the management competency rating model and the impact of the individual input variable to the management competency rating model.

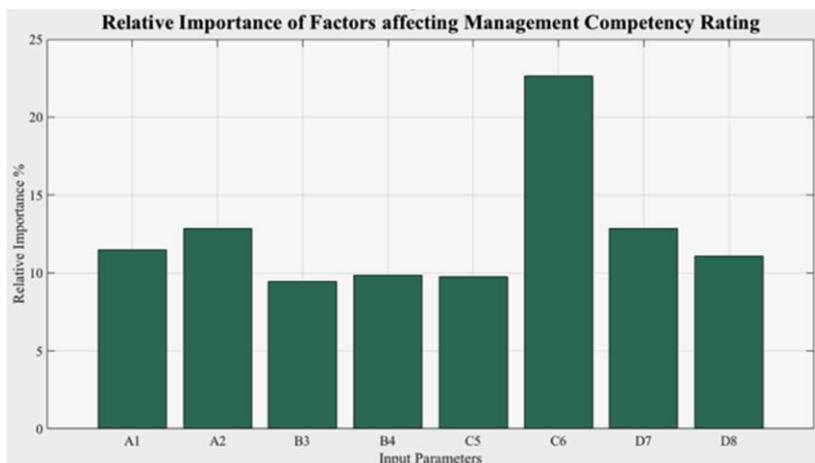


**Figure 5.** Correlation of actual vs. Predicted management competency rating.

**Table 2.** Product weights of the management competency rating model.

	A1	A2	B3	B4	C5	C6	D7	D8
Hidden 1	0.1807	0.1852	0.0707	0.0638	0.0258	0.3753	0.0530	0.0455
Hidden 2	0.0377	0.2255	0.0895	0.0281	0.0520	0.3129	0.2408	0.0136
Hidden 3	0.1802	0.0762	0.1292	0.1045	0.2299	0.1546	0.0241	0.1012
Hidden 4	0.0845	0.0076	0.1517	0.2359	0.0647	0.1391	0.1097	0.2069
Hidden 5	0.0817	0.1373	0.0100	0.1896	0.2711	0.1703	0.0224	0.1175
Hidden 6	0.1372	0.0515	0.0443	0.0120	0.0072	0.2685	0.2368	0.2424
Hidden 7	0.1060	0.1661	0.1411	0.0983	0.1197	0.1962	0.1704	0.0021
Hidden 8	0.1118	0.1784	0.1186	0.0556	0.0120	0.1953	0.1709	0.1575
Sum	0.9198	1.0279	0.7552	0.7877	0.7823	1.8122	1.0281	0.8868
<b>R.I.</b>	<b>11.497</b>	<b>12.849</b>	<b>9.440</b>	<b>9.846</b>	<b>9.779</b>	<b>22.653</b>	<b>12.852</b>	<b>11.085</b>

The ranking of influence of the input parameters to the management competency rating were as follows: B3<C5<B4<D8<A1<A2<D7<C6. The C6 – Managing Innovation is the most influential parameter to the management competency while the B3 – people management and authentic leadership is the least influential parameter based on the respondents answer to the survey questionnaire tool as shown in Figure 6.

**Figure 6.** Relative Importance of Factors affecting Management Competency.

#### 4. Conclusion

This study focused on developing a neural network model for predicting the management competency rating and the relative importance of factors affecting it. The study presents that the predicted results of the developed neural network model showed an excellent performance with high  $R_{all}$  and very low MSE and MAPE. It has final characteristics of 8-8-1 (input-hidden-output). The sensitivity analysis showed that the ranking of importance for the factors affecting management competency rating to be B3<C5<B4<D8<A1<A2<D7<C6. Overall, the neural network model created in this research yielded an excellent and accurate results in predicting the management

competency rating as well as using its connection weights to calculate the influence of each input parameters to the management competency rating.

## Acknowledgments

This paper acknowledges Mapua University for the support vital for the fulfillment of this research.

## References

- [1] Edum-Fotwe FT, McCaffer R. Developing project management competency: perspectives from the construction industry. *Int J Proj Manag.* 2000; 18(2): 111 – 24.
- [2] Silva D, Villaverde B, De Jesus KL, Marcial Jr, Ejera RL, Villa-Real CV, Zarrage JM. Design initiative implementation framework: A model integrating kolmogorov-smirnov in sustainable practices for triple-bottom-line principles in construction industry. *Civ Eng Archit.* 2020; 8(4): 599 – 617.
- [3] Silva DL, Sabino LD, Lanuza DM, Adina EM, Villaverde BS, Pena EG. Silva's management competency theory: a factor-item analytic approach utilizing oblique rotation direct oblimin method under Kaiser-Bartlett's test of sphericity. In Proceedings of the World Congress on Engineering and Computer Science 2014 Vol. I WCECS 2014; 2014 Oct 22; San Francisco, CA. Vol. 24.
- [4] Kulkarni P, Londhe S, Deo M. Artificial neural networks for construction management: a review. *J Soft Comput Civ Eng.* 2017; 1(2): 70 – 88.
- [5] Patel DA, Jha KN. Neural network model for the prediction of safe work behavior in construction projects. *J Constr Eng Manag.* 2015; 141(1): 04014066.
- [6] Heravi G, Eslamdoost E. Applying artificial neural networks for measuring and predicting construction-labor productivity. *J Constr Eng Manag.* 2015; 141(10): 04015032.
- [7] Wang D, Ardit D, Damci A. Construction project managers' motivators and human values. *J Constr Eng Manag.* 2017; 143(4): 04016115.
- [8] Khanzadi M, Nasirzadeh F, Dashti MS. Fuzzy cognitive map approach to analyze causes of change orders in construction projects. *J Constr Eng Manag.* 2018; 144(2): 04017111.
- [9] Kim GH, Seo DS, Kang KI. Hybrid models of neural networks and genetic algorithms for predicting preliminary cost estimates. *J Comput Civ Eng.* 2005; 19(2): 208 – 11.
- [10] Oreta AW, Kawashima K. Neural network modeling of confined compressive strength and strain of circular concrete columns. *J Struct Eng.* 2003; 129(4): 554 – 61.
- [11] Ullah S, Tanyu BF, Zainab B. Development of an artificial neural network (ANN)-based model to predict permanent deformation of base course containing reclaimed asphalt pavement (RAP). *Road Mater Pavement Des.* 2020; 1 – 19.
- [12] Rinchon JPM, Concha NC, Calilung MGV. Reinforced concrete ultimate bond strength model using hybrid neural network-genetic algorithm. In 2017 IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM); 2017; 1 – 6.
- [13] Silva DL, de Jesus KLM, Villaverde BS, Adina EM. Hybrid Artificial Neural Network and Genetic Algorithm Model for Multi-Objective Strength Optimization of Concrete with Surkhi and Buntal Fiber. In Proceedings of the 2020 12th International Conference on Computer and Automation Engineering; 2020; 47 – 51.
- [14] Henríquez PA, Ruz GA. A non-iterative method for pruning hidden neurons in neural networks with random weights. *Appl Soft Comput.* 2018; 70: 1109 – 1121.
- [15] Mountford GL, Atkinson PM, Dash J, Lankester T, Hubbard S. Sensitivity of Vegetation Phenological Parameters: From Satellite Sensors to Spatial Resolution and Temporal Compositing Period. In *Sensitivity Analysis in Earth Observation Modelling*; 2017; Elsevier; 75 – 90.
- [16] Laroza Silva D, Marcelo De Jesus KL. Backpropagation Neural Network with Feature Sensitivity Analysis: Pothole Prediction Model for Flexible Pavements using Traffic and Climate Associated Factors. In 2020 the 3rd International Conference on Computing and Big Data; 2020; New York, NY, USA: Association for Computing Machinery, 60 – 67.

# Hybrid Neural – Based Model for Predicting the Construction Project Performance of High – Rise Building Projects in Metro Manila, Philippines

Jucar FERNANDEZ<sup>a</sup>, Dante L. SILVA<sup>a,b1</sup>, Kevin Lawrence M. De JESUS<sup>a</sup>

<sup>a</sup>*School of Graduate Studies, Mapua University, Manila, Philippines*

<sup>b</sup>*School of Civil, Environmental and Geological Engineering, Mapua University,  
Manila, Philippines*

**Abstract.** Assessing construction project performance through adapting an innovative approach can multiply the production of high-quality project outputs which is a catalyst to the socio-economic progress of a country. Preliminary data collection was employed through a meta-cognitive analysis of reviewing related literature which directs to the backbone of qualitative information that is relevant to the periodically experienced construction performance-influencing factors and to develop an assessment questionnaire about the influencing factors affecting the project performance. The IBM SPSS program was used to verify the reliability and consistency of the fundamental statistics of the questionnaire responses of cost, time and quality performance ratings. A predictive mathematical model was developed for forecasting cost, time, and quality performance rating employing Levenberg-Marquardt training algorithm with Hyperbolic Tangent Sigmoid function. The prediction model result shows a highly satisfying performance on its variance from the substantive values and suggests a high correlation between these values. The relative importance of the factors affecting the cost, time, and quality performance rating was calculated via sensitivity analysis through connection weights using Garson's Algorithm to view the order of influence of the parameters that have great significance to the success of a project.

**Keywords:** Construction analytics, project performance, neural network, sensitivity analysis, KPI dashboard

## 1. Introduction

Construction is considered a major industry throughout the world which calculating a colossal proportion of most countries' gross domestic product and gross national product. The key importance of the construction sector is not only with its whopping size but also its eminent role as a catalyst in economic growth. Assessing construction project performance using a cutting-edge approach can multiply the production of high-quality project outputs which is a critical contributor to the social and economic development of a country [1].

---

<sup>1</sup>Corresponding author: Dante L. SILVA, School of Graduate Studies, Mapua University, Manila, Philippines; E-mail: dantesilva2000@yahoo.com.

The construction industry is very complex because it involves different parties of internal and external stakeholders: owners, consultants, general contractors, and subcontractors [2]. Moreover, several Key Performance Indicators (KPI) had been employed and used to assess and evaluate project performance. However, the time, cost, and quality indicators are the most prevalent performance evaluation factors which widely used on construction analytics and reporting dashboards [3]. Acknowledging the impact of a construction project in this time of pandemic for stakeholder key goals for the overall growth of the project suggests that there is a compelling need to ensure the quality of management tools and approach throughout the execution process. Therefore, it is vital nowadays to develop a new technique and adopt innovation to provide solutions to different problems faced by projects related to time, cost, and quality factors. Using project performance, measures, and assess the project and organizational performance in the entire project life cycle [4].

Artificial Intelligence (AI) has been employed in the construction industry for the past decade [5]. Moreover, the utilization of breakthrough tools such subfield of Artificial Intelligence Technique was in the position of creating a predictive model and in lieu of the progression of different practices to 4<sup>th</sup> Industrial Revolution and can evaluate data sets with sequential and complex relationship utilizing a hybrid artificial neural network [6].

The principal goal of this study is to formulate a critical key indicator framework employing a construction analytics and reporting dashboard with a hybrid neural-based model for predicting the construction project performance of high-rise building projects in Metro Manila, Philippines.

## 2. Methodology

### 2.1. Meta – Cognitive Analysis from Reviewing Literature and Conducting Surveys and Interviews with Experts

Substantial number of studies has been utilized to investigate the factors influencing the project performance in different countries [7]. Through the meta-cognitive analysis from reviewing the literature of previous studies, many factors affecting the project performance rating were listed down. Then, 5 local experts in construction supervision, construction management, and consulting field have been interviewed and surveyed to help in developing the final list of influencing factors that will part of the assessment questionnaire. There are 18 influencing factors that were considered predominant factors for the project performance in correlation to cost, time, and quality indicators which is presented in table 1.

**Table 1.** Factors affecting Cost Performance (CP), Time Performance (TP) and Quality Performance (QP) Rating.

Factor	Notation
The total floor area of the building	CP1/TP1/QP1
Issues arise on the project site	CP2/TP2/QP2
Completeness of the contractor's plant and equipment	CP3/TP3/QP3
Financial capacities of the general contractor	CP4/TP4/QP4
Technical capacities of contractor's personnel	CP5/TP5/QP5
Accessibility of trained resources	CP6/TP6/QP6
Favourable working condition	CP7/TP7/QP7
Nature of building	CP8/TP8/QP8
Contractor's performance of achieving project finish date milestone	CP9/TP9/QP9
Contractor's performance of achieving quality work	CP10/TP10/QP10
Competency level of contractor's staff	CP11/TP11/QP11
Level of occupational specialty requisite of contractors	CP12/TP12/QP12
Subcontractor's knowledge and means	CP13/TP13/QP13
Contractor's cost expenditure performance	CP14/TP14/QP14
Design capacities when the budget is set	CP15/TP15/QP15
Level of design complexity	CP16/TP16/QP16
Climatic conditions at the project site	CP17/TP17/QP17
Level of construction complexity	CP18/TP18/QP18

## 2.2. Neural Network Modelling

Levenberg – Marquardt (*trainlm*) for training algorithm and Hyperbolic Tangent Sigmoid (*tansig*) for the transfer function was used in the development of the predictive mathematical model [8]. Furthermore, Levenberg – Marquardt emerges as the most advanced approach for medium – sized feed – forward ANN, and it may be implemented successfully to several hundred weights [9].

For this study, 18 hidden neurons and 2 hidden layers were used. Pearson's Correlation Coefficient (R), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE) were used as performance indicators [10]. The created model's efficacy and good level performance for cost, time and quality performance rating must have extremely high R values and extremely low MSE and MAPE values [11].

## 2.3. Sensitivity Analysis

A sensitivity analysis was implemented in this study through connection weights Garson's algorithm that was demonstrated to view the ranking of influence of the lists of factors of each performance rating [12].

## 3. Results and Discussion

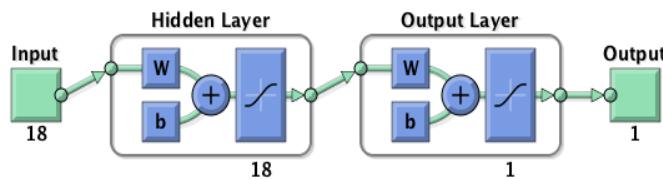
### 3.1. Neural Network Model Simulation Results

The cost, time, and quality performance rating models have internal characteristics presented in Table 2. The training algorithm employed in model development was the Levenberg-tansig function. The number of hidden neurons used for the performance model development is equal to the Marquardt algorithm while the transfer function used

is the tansig with “m” number of hidden neurons [13]. The design of the developed performance rating models is shown in Figure 1.

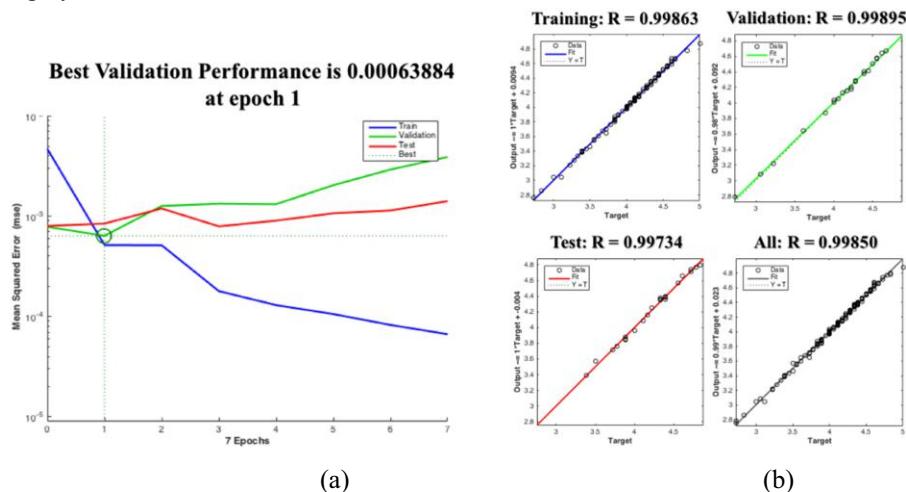
**Table 2.** Description of the details of the cost, time and quality performance rating models.

Internal Characteristics	Value
Training Algorithm	trainlm
Transfer Function	tansig
Number of Hidden Neurons	18
Number of Hidden Layers	2 (1 Hidden Layer and 1 Output Layer)
Performance Indicators	R, MSE and MAPE

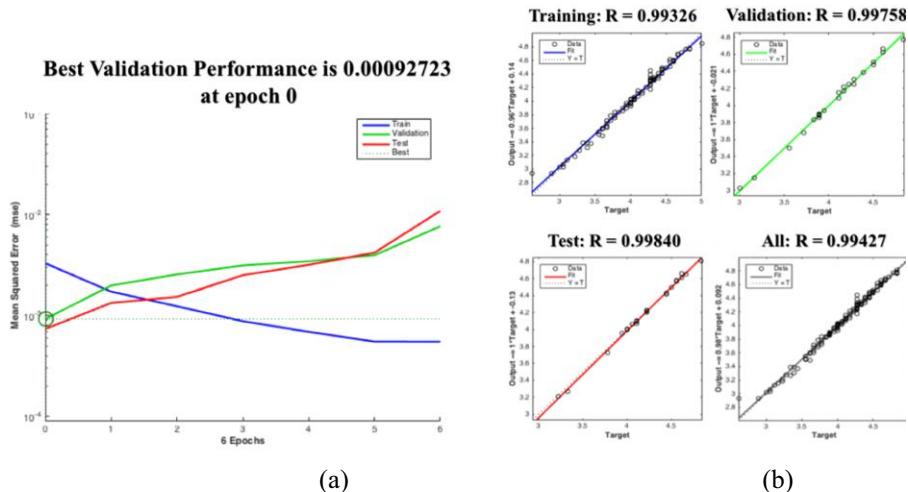


**Figure 1.** Design of the Performance Rating Models.

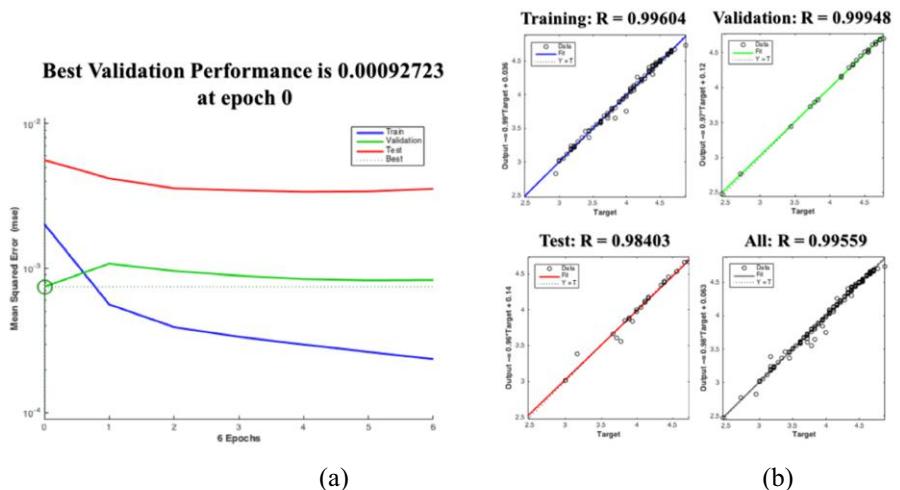
As demonstrated in Figure 2 to Figure 4, the performance rating models had a very high  $R_{all}$  value and a very low MSE and MAPE. The connection between the actual and anticipated performance ratings shows that the actual and predicted performance ratings are highly correlated.



**Figure 2.** (a) Performance Plot and (b) Correlation Plots of the Cost Performance Rating Model.



**Figure 3.** (a) Performance Plot and (b) Correlation Plots of the Time Performance Rating Model.



**Figure 4.** (a) Performance Plot and (b) Correlation Plots of the Quality Performance Rating Model.

### 3.2. Variable Significance of the Parameters Influencing Cost, Time and Quality Performance Rating

The variable significance of the factors affecting cost, time, and quality was calculated utilizing the connection weights of the network using Garson's Algorithm. Figure 5a presents the variable significance of the factors affecting cost performance rating. The ranking of influence of these factors is described as CP18<CP4<CP17<CP13<CP11<CP15<CP12<CP8<CP2<CP9<CP6<CP3<CP1<CP10 <CP16<CP14<CP7<CP5 wherein CP18 – a level of construction complexity is the least influential factor with R.I. = 4.13200% and CP5 – technical capacities of contractor's personnel are the most influential factor with RI = 6.89412%.

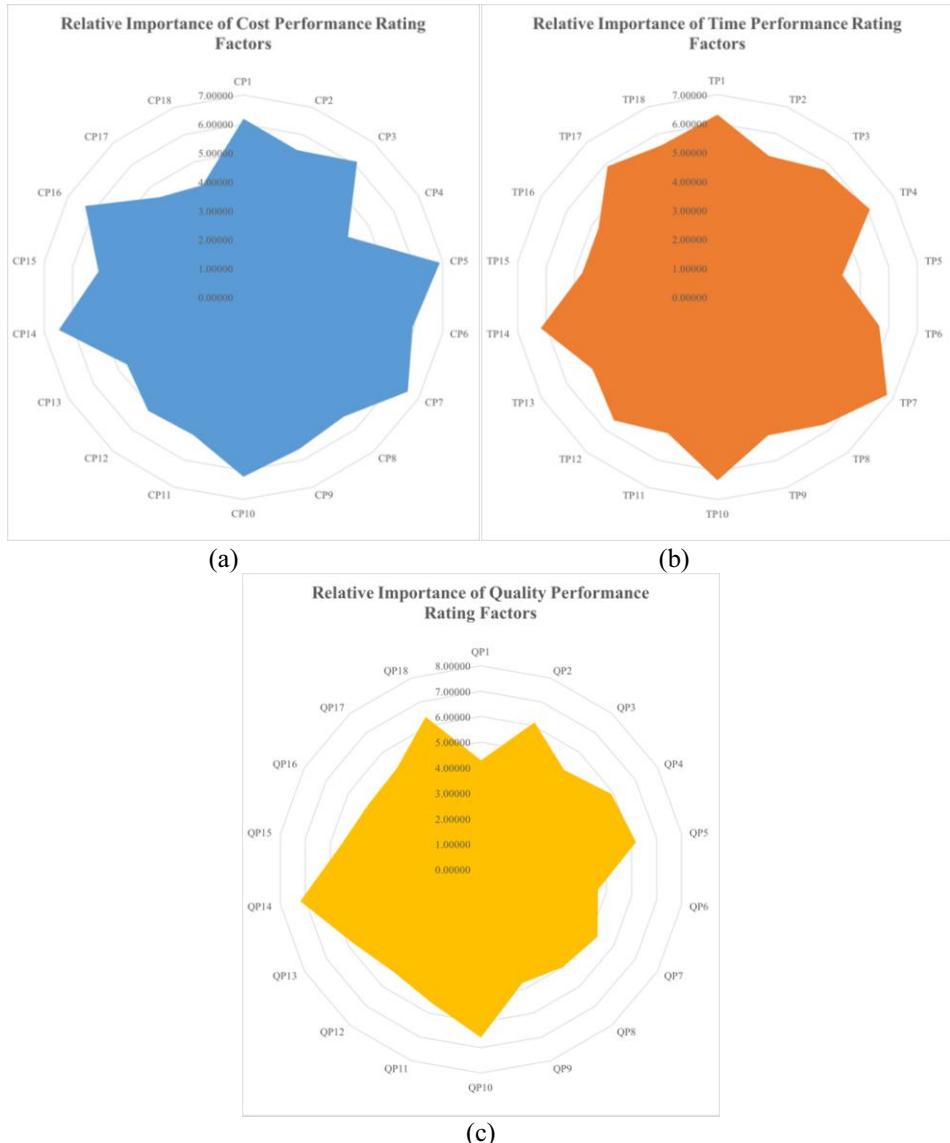


Figure 5. (a) Performance Plot and (b) Correlation Plots of the Quality Performance Rating Model.

Figure 5b presents variable significance of the factors affecting time performance rating. The ranking of influence of these factors is described as  $TP5 < TP15 < TP16 < TP13 < TP11 < TP9 < TP2 < TP12 < TP18 < TP6 < TP8 < TP3 < TP17 < TP4 < TP14 < TP1 < TP10 < TP7$  wherein TP5 – technical capacities of contractor's personnel is the least influential factor with R.I. = 4.37899% and TP7 – favorable working conditions is the most influential factor with RI = 6.76326%.

Figure 5c present the product weights and variable significance of the factors affecting quality performance rating. The ranking of influence of these factors is described as

QP1<QP6<QP9<QP8<QP3<QP16<QP17<QP7<QP12<QP15<QP11<QP13<QP4<QP2<QP5<QP18<QP10<QP14 wherein QP1 – total floor area of the building is the least influential factor with R.I. = 4.27745% and QP14 – contractor's cost expenditure performance is the most influential factor with RI = 7.20907%.

### 3.3. KPI Dashboard

As the actual data gathered and validated by the staff of the project with the approval of the site management, this engineering tool able to provide a solution to unseen and overlook problems and issues faced by the project. With this tool, easy access to the status of the late response of the designer and the project management on the review of submittals and the pending raise claims as events that will part in the substantiation for the claim. Furthermore, the productivity output of the manpower can be seen and evaluated instantly and assess its result from the standard requirement of the project. Lastly, with the help of this KPI Dashboard shown in Figure 6, the site management, and the high management of a local company able to picture-out the real-time condition of the project.

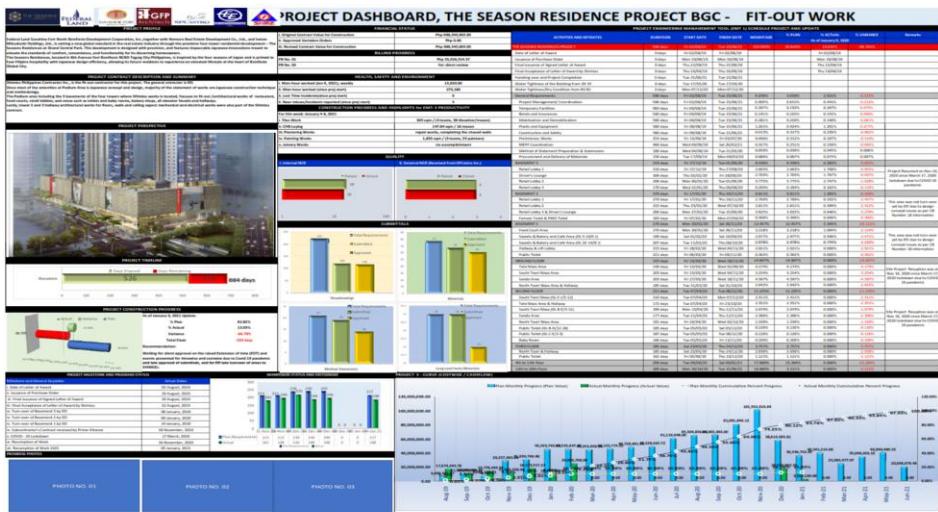


Figure 6. Developed KPI Dashboard.

#### 4. Conclusion

The created performance rating models for cost, time and quality has exceptionally high R values and extremely low MSE and MAPE values, indicating the proposed model's efficacy and performance.

Garson's algorithm was used to calculate the variable significance of the factors for performance rating of time, cost, and quality. The result is close to the output of actual validation and evaluation conducted on one of an on-going high-rise building project in Manila, Philippines. This is as per their actual progress on the different performance indicators using construction reporting dashboards.

To sum up, project performance can be measured, and the output presented on the training and validation test of one high-rise building projects show that the developed model can be a guide for on-going and future projects. Thus, the artificial neural network is suitable and effective for predicting construction project performance.

#### References

- [1] Cabuñas J, Silva D. Exploratory factor – Item analytic approach for construction project cost overrun using oblique promax rotation for predictors determination. *Int J Innov Technol Explor Eng.* 2019; 8: 2278 – 3075.
- [2] Alias M, Dhanya R, Ramasamy G. Study on factors affecting the performance of construction projects and developing a cost prediction model using ANN. *Int J. Earth Sci. Eng.* 2015; 8(5): 2189 – 2194.
- [3] Mossalam A, Arafa M. Using artificial neural networks (ANN) in projects monitoring dashboard's formulation, housing, and building national research center. *HBRC J.* 2018; 14(3): 385 – 392.
- [4] Omar MN, Fayek AR. Modeling and evaluating construction project competencies and their relationship to project performance *Autom Constr.* 2016; 69: 115 – 130.
- [5] Golnaraghi S, Zangenehmadar Z, Moselhi O, Alkass S. Application of artificial neural network (s) in predicting formwork labour productivity. *Adv Civ Eng* 2019. 2019.
- [6] Yu THK, Huarng KH. A neural network-based fuzzy time-series model to improve forecasting. *Expert Syst Appl.* 2010; 37(4): 3366 – 3372.
- [7] Jha K. Construction project management: Theory and practice. India: Pearson Education; 2011.
- [8] Laroza Silva D, Marcelo De Jesus KL. Backpropagation Neural Network with Feature Sensitivity Analysis: Pothole Prediction Model for Flexible Pavements using Traffic and Climate Associated Factors. In 2020 the 3rd International Conference on Computing and Big Data; 2020; New York, NY, USA: Association for Computing Machinery, 60 – 67.
- [9] Hagan MT, Menhaj MB. Training feedforward networks with the Marquardt Algorithm. *IEEE trans neural netw.* 1994; 5(6): 989 – 993.
- [10] Anupama S. Prediction of construction project performance using regression analysis and artificial neural network. *Int J Res Appl Sci Technol.* 2018; 6(5): 1172 – 1179.
- [11] Silva DL, de Jesus KLM, Villaverde BS, Adina EM. Hybrid Artificial Neural Network and Genetic Algorithm Model for Multi-Objective Strength Optimization of Concrete with Surkhi and Buntal Fiber. In Proceedings of the 2020 12th International Conference on Computer and Automation Engineering; 2020; 47 – 51.
- [12] Olden JD, Joy MK, Death RG. An accurate comparison of methods for quantifying variable importance in artificial neural networks simulation data. *Ecol model.* 2004; 178(3-4): 389 – 93.
- [13] Kim D, Kim H, Chung D. A modified genetic algorithm for fast training neural networks. In International Symposium on Neural Networks; 2005 May; Berlin, Heidelberg: Springer; 660 – 665.

# Simulation of Deformation Transfer Coefficient of Pipe Bend Buried Based on Shaking Table Test and Goodman Contact Element

Delong HUANG<sup>a1</sup>, XiaoDong BAI<sup>b</sup>, Xiaoli CHEN<sup>b</sup>, Guanyu XU<sup>b</sup>, Aiping TANG<sup>a</sup>

<sup>a</sup>*School of Civil Engineering, Harbin Institute of Technology, Harbin, China*

<sup>b</sup>*Department of Architecture, Heilongjiang College of Construction, Harbin, China*

**Abstract.** Due to the large difference of stiffness between pipe and soil, the movement of the two can not be coordinated under seismic. Therefore, the deformation transfer between pipe and soil is a very important research content in the study of pipe failure. At present, scholars have done less research on the pipe-soil deformation transfer of elbow. In this paper, the fitting formula of deformation transfer coefficient of buried elbow under seismic action was obtained by scale shaking table test of pipe bend and 3D finite element model based on Goodman contact element. Then, the test results are compared with the calculation results of the fitting formula and the simulation results of the finite element method to verify the rationality of the fitting formula and analyze the change law of the deformation transfer coefficient at the elbow of the pipe, including the influence of different pipe diameters, buried depth, wall thickness, soil properties, and elbow angles. It is confirmed that these factors have a great influence on the deformation transfer between the pipe and soil, which indicates that the fitting formula of the deformation transfer coefficient at the elbow is of huge significance to the earthquake resisting design of pipe.

**Keywords:** Deformation transfer coefficient, elbow, shaking table test, goodman contact element, interaction of pipe and soil, earthquake resisting design of pipe

## 1. Introduction

Water supply and gas supply pipeline networks are an important part of the city's lifeline. In many seismic investigations, it is found that the damage of underground pipe network may damage water supply pipelines, fire-fighting devices and natural gas facilities. Moreover, earthquakes may also cause urban drainage problems and many secondary disasters, and bring great disasters to human life and production [1]. Hanshin Earthquake ( $M=7.3$ ), more than 20,000 natural gas pipelines leak occurred in Osaka, most of which occurred in places where rivers are prone to liquefaction and welded steel pipe elbows in the 1995. In terms of water supply pipelines, only 900,000 households in Kobe city had no water supply, 23 elbows near the Shenqi River were damaged, and the sewage and rainwater discharge pipeline network in Kobe City was also severely damaged,

---

<sup>1</sup>Corresponding author: Delong Huang, School of Civil Engineering, Harbin Institute of Technology, Harbin, China; E-mail: huang06080601@163.com.

reaching nearly 3m/ha [2]. In the 2008 Wenchuan earthquake ( $M=8.0$ ), water supply and gas supply facilities including underground pipe network in nearly 20 cities and counties in Sichuan Province were seriously damaged. For example, the gas facilities and drinking water companies in Mianzhu, Beichuan, Dujiangyan and other cities have not been damaged, while the buried water and gas supply networks have been damaged, and the connections between pipelines have been seriously damaged, which can not be used normally and is difficult to repair [3]. Therefore, in the network design, the pipeline system in the city is the key facility to ensure people's livelihood.

In the earthquake, for the simple pipe-soil interaction, the buried pipeline will be damaged by the force generated by the pipe-soil interface interaction, in which the ground movement, wave propagation or permanent ground displacement (PGD) will cause damage to the buried pipeline [4]. The permanent ground deformation includes landslide, fault, lateral expansion and settlement, and the movement is not coordinated due to the deformation of the soil, resulting in uneven deformation of local surface of the pipeline. Therefore, the pipe-soil deformation transfer has a greater impact on the narrow failure of the pipeline network [5]. The local damage of the pipe network, such as the damage to the elbow and the damage of the pipe network nodes, is bound to bring greater challenges to the repair of the whole pipe network.

Currently, scholars' research on underground pipeline bending is limited to two aspects. The first is the reliability for network topology and network connection in disaster situations. These studies were based on the idea of probability, only pay attention to the damage probability of underground pipeline network, but do not study the mechanical mechanism of elbow damage [6-9]. Other aspect is the research on the safe operation and risk assessment of the pipeline network under earthquake conditions. This aspect is the deduction of the algorithm based on the monitoring data. These studies are the evaluation of the seismic performance of the existing underground pipeline network and the evaluation of the safe operation of the pipeline network through the data fusion of sensors, but they are not involved in the mechanical failure model of the pipeline network [10-19]. From the above analysis, it can be observed that many scholars no longer study single pipe, but study the underground pipe network involving bends. Their research is aimed at the probability analysis and algorithm research for pipelines, and the mechanical mechanism of bend failure is not mature.

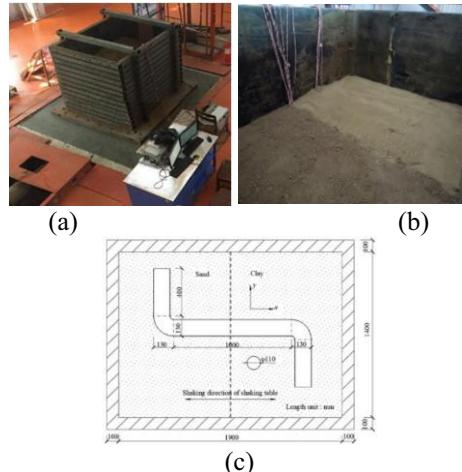
For this study, shaking table test and finite element simulation based on Goodman contact element were utilized to study the coefficient of deformation transfer under seismic, and the formula was fitted according to the influencing factors. Then, change law of deformation transfer coefficient was discussed through comparison, which provided a theoretical basis for design of the bend.

## 2. Design of the Test

Due to the limitation of existing conditions, PVC-U pipe (unplasticized polyvinyl chloride pipe) was used to simulate iron pipe. PVC-U pipe is a rigid plastic pipe. According to the influence of model scale, the pipeline is not easy to be damaged by seismic wave and is always in the elastic stage. Therefore, the internal force and deformation trend of PVC-U pipe areas of iron pipe. Through the similarity of elastic modulus, PVC-U pipe can be utilized to simulate iron pipe. Figure 1 (c) below shows the form, layout, and dimensions of the elbow. The internal dimension of the shear box is  $1900 \times 1400\text{mm}$ , with a diameter of 100mm (75mm and 50mm in comparison test), a

wall thickness  $t$  of 2.0mm (3.0mm and 4.0 mm in comparison test), an elastic modulus of 4000 MPa, a density of 1400 kg/m<sup>3</sup>, a Poisson's ratio of 0.3 and a buried depth of 0.3 m (0.2 m and 0.4 m in the comparative test).The elbow angle is 90 degrees (45 degrees and 135 degrees in the comparative test). The elbow was inserted into the straight pipe and coated with PVC-U special glue. Because the bonding strength of the special glue can reach the strength of the pipe itself, this form can simulate the welding of iron elbow. In the test, the end of the pipe was sealed with rubber to prevent the soil from entering the pipe and weaken the influence of shear box.

In this test, sand and clay were loaded into the laminated shear box on both sides. The shear box as shown in Figure 1(a) and compacted layer by layer as shown in Figure 1(b). The compacted height of the soil is 15cm for each layer. This test uses the volume and quality of each layer of soil to control the density of the filled soil. Therefore, it is slightly soft for the soil, according to the state, the shear wave velocity of the soil is about 120 m/s. After filling and tamping, a soil sample is collected from the shear box, and the physical parameters of the soil are measured with a pycnometer and a ring knife. For sand, the following physical parameter values were measured: density  $\rho = 1692 \text{ kg/m}^3$ , specific gravity  $G_s = 2.62$ , the internal friction angle  $\varphi = 32^\circ$ , cohesion  $c = 0$  and water content  $\omega = 0.18\%$ . The corresponding values for the clay were: density  $\rho = 1430 \text{ kg/m}^3$ , specific gravity  $G_s = 2.45$ , the angle of internal friction  $\varphi = 20^\circ$ , cohesion  $c = 5 \text{ kPa}$ , and water content  $\omega = 10.65\%$ .

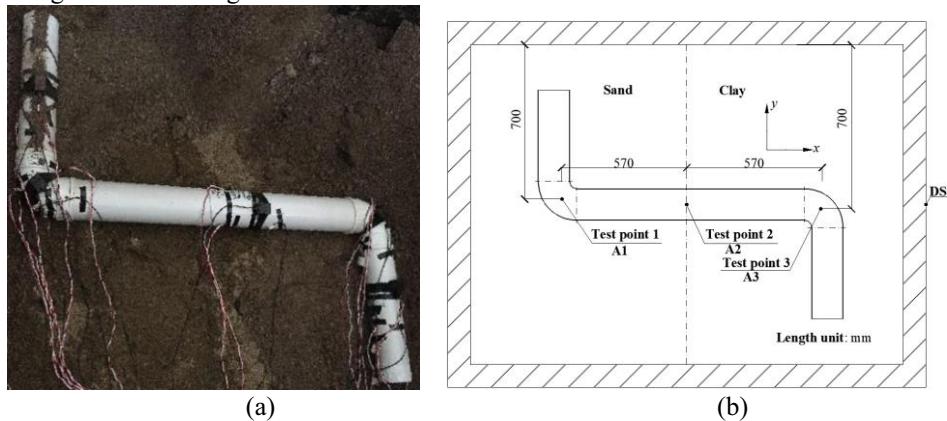


**Figure 1.** Design of test shear box, soil and pipe bend. (a) Laminated shear box. (b) The division of sand and clay. (c) Schematic diagram of pipe laying.

The acceleration measurement adopted the acceleration sensors that were produced at B&K in Denmark, with a measurement accuracy of 0.1 m/s<sup>2</sup>. The acceleration sensor was installed in a closed iron box and fixed to the pipe to prevent interference from the outside soil, as shown in Figure 2(a). The arrangement of the accelerometers has been shown at A1, A2, and A3 in Figure 2(b), which were used to measure the acceleration of the first (Test point 1), second (Test point 2) and third section (Test point 3). Test point 1 and Test point 3 were located at the elbow. A DH5922N data acquisition system from the Donghua test company was adopted for data acquisition, with a sampling frequency of 1000 Hz.

The displacement sensor used in this test is LVDT (linear variable differential transformer), and its measurement accuracy is 0.1mm. Since the sensor measures the distance from the solid surface through expansion and contraction with the solid surface, it can only be arranged on the outer surface of the solid. It is arranged on the outer surface of the shear box and at the same height as the pipe, represented by DS. Because the size of the soil box is small, the displacement of the soil around the pipeline is approximately equivalent to that of the shear box, as shown in Figure 2 (b).

This experiment used dimensional analysis to identify similar relationships. Length  $l$ , elastic modulus  $E$ , and acceleration  $a$  were basic physical quantities. In multi-media coupling dynamic test, the similar relationship between the foundation soil and the pipeline structure should be kept as consistent as possible, and the soil should be tamped to be consistent with the actual site as far as possible. For the shaking table test, the similar design was mainly pipe structure. In the design of similarity ratio, the table size, dynamic performance, load-bearing tonnage and other supporting equipment performance of the test system should need to be fully considered. The similarity ratio design of this test is given in Table 1.



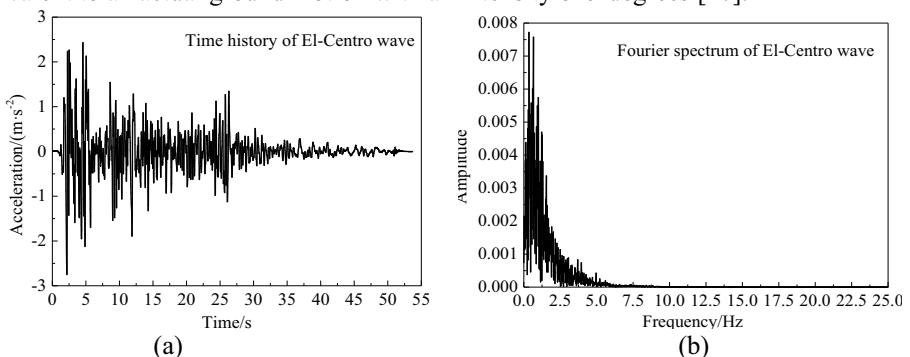
**Figure 2.** Sensors layout for pipe bend. (a) The sensors layout at pipe bend. (b) The design of sensors layout for pipe bend.

**Table 1.** Similarity ratio design of shaking table test.

Groups	Physical parameters	Similar relations	Similar ratio
<b>Geometric parameters</b>	Length $l$	$S_l$	0.10
	Area $S$	$S_S = S_l^2$	0.01
	Linear displacement $u$	$S_u = S_l$	0.10
	Angular displacement $\theta$	$S_\theta = S_\sigma / S_E$	1.00
<b>Material properties</b>	Elastic Modulus $E$	$S_E$	0.02
	Stress $\sigma$	$S_\sigma = S_E$	0.02
	Strain $\epsilon$	$S_\epsilon = 1$	1.00
	density $\rho$	$S_\rho = S_E / (S_l S_a)$	0.10

Load performance	quality $m$	$S_m = S_\sigma S_l^2 / S_a$	1.00×10 <sup>-4</sup>
	Concentrated force $F$	$S_F = S_E S_l^2$	2.00×10 <sup>-4</sup>
	Linear load $q$	$S_q = S_E S_l$	2.00×10 <sup>-3</sup>
	Surface load $p$	$S_p = S_E$	0.02
	Moment $M$	$S_M = S_E S_l^3$	2.00×10 <sup>-5</sup>
	Time $t$	$S_t = S_l^{0.5} S_a^{-0.5}$	0.22
	Frequency $f$	$S_f = S_l^{-0.5} S_a^{0.5}$	4.47
Dynamic characteristics	Velocity $V$	$S_v = S_l^{0.5} S_a^{0.5}$	0.45
	Damping $C$	$S_c = S_\sigma S_l^{1.5} S_a^{-0.5}$	4.47×10 <sup>-4</sup>
	Acceleration $a$	$S_a$	2.00

Due to the limitations of the test equipment, when seismic waves were loaded, the seismic station only moves horizontally in one dimension. Consider the most unfavorable seismic wave input direction when designing the experiment. This paper uses El-Centro wave as the input prototype wave in Figure 3(a). The predominant frequency of Fourier spectrum corresponding to seismic wave is 1Hz in Figure 3(b). Because of the similarity ratio of the test, the seismic wave shape input by the shaking table must be tailored according to the similarity condition, that is, the prototype wave must be adjusted according to the time similarity ratio. In the shaking table test for bend in this paper, the pipeline moves along the  $x$  direction of Figure 1(c), and its peak value was converted to 0.8 g according to the similarity relationship. Intensity of this ground motion is equivalent to an actual ground motion with an intensity of 9 degrees [20].



**Figure 3.** Prototype curves of seismic wave loaded. (a) The curve of time history. (b) The curve of Fourier spectrum.

### 3. Test Results and Formula of Deformation Transfer Coefficient

#### 3.1. Test Results of the Pipe Bend

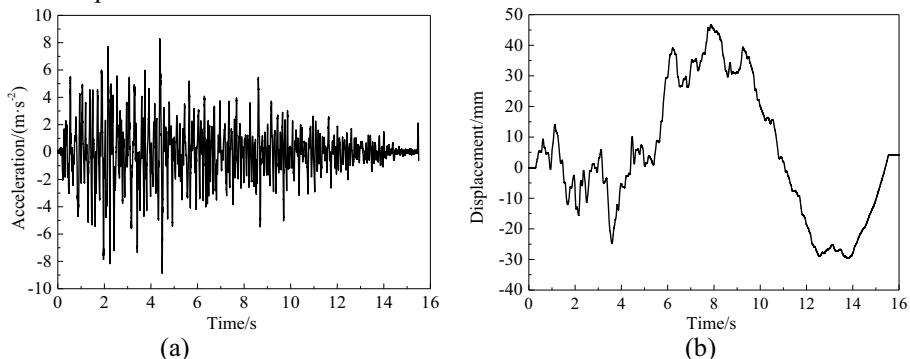
The acceleration of the pipe could be measured by the acceleration sensor A1 ( $D = 0.1$  m,  $H = 0.3$  m,  $t = 2.0$  mm,  $\theta = 90^\circ$ ) in the soil box of the site shaking table, as shown

in Figure 4(a), and the displacement could be obtained by integrating twice, seeing Figure 4(b). The maximum displacement was 45.912 mm, and the corresponding time was 7.875s. When loading, the maximum value for the sensor DS measuring the soil displacement was 140.607 mm, and the corresponding time of the maximum displacement was 7.876 s, which indicated that the time of the maximum pipe-soil displacement was almost the same, as shown in Figure 5.

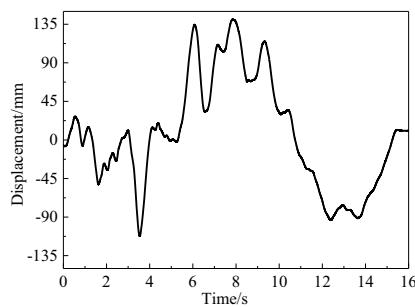
With the above method, the measured accelerations of A1 and A3 under different working conditions of four factors were integrated from 0 to 16.0 s, the displacement of the first 16.0 s was obtained, and the actual displacement of Test points 1 and Test point 3 was obtained, and the maximum displacement was found out. Then found out the maximum value of the soil displacement from the data recorded by the displacement sensor corresponding to each working condition, and utilized equation (1) to calculate the deformation transfer coefficient [21] of the elbow under each working condition, as shown in Table 2.

$$\beta = \frac{\Delta_p}{\Delta_s} \quad (1)$$

where  $\Delta_p$  is the actual absolute displacement of the elbow;  $\Delta_s$  is the actual absolute displacement of the soil.



**Figure 4.** Time history curve of acceleration A1 and displacement curve obtained by integration. (a) Acceleration A1 time history. (b) Displacement integral time history.



**Figure 5.** Displacement curve of soil around elbow with time.

**Table 2.** Comparative test results of displacement and pipe-soil deformation transfer coefficient.

Pipe diameter /m	Buried depth /m	Wall thickness /mm	Angle /°	Pipe displacement /cm		Soil displacement /cm	Deformation transfer coefficient	
				Sand	Clay		Sand	Clay
0.050	0.3	2.0	90	5.53	6.63	11.26	0.49	0.59
0.075	0.3	2.0	90	5.66	6.89	13.70	0.41	0.50
0.100	0.3	2.0	90	4.59	6.39	14.06	0.33	0.45
0.100	0.2	2.0	90	5.08	6.81	17.38	0.29	0.39
0.100	0.3	2.0	90	4.59	6.39	14.06	0.33	0.45
0.100	0.4	2.0	90	5.06	6.41	12.58	0.40	0.51
0.100	0.3	2.0	90	4.59	6.39	14.06	0.33	0.45
0.100	0.3	3.0	90	3.56	4.42	13.03	0.27	0.34
0.100	0.3	4.0	90	2.83	3.40	13.92	0.20	0.24
0.100	0.3	2.0	45	5.12	6.92	11.60	0.44	0.60
0.100	0.3	2.0	90	4.59	6.39	14.06	0.33	0.45
0.100	0.3	2.0	135	2.50	3.79	12.79	0.20	0.30

Note: at this time, the seismic wavelength is 2 m; the elastic modulus of the pipe is 4000 MPa; the foundation coefficient is 5 MPa for clay and 3 MPa for sand; the pipe-soil friction coefficient is 0.4.

### 3.2. Formula Fitting of Pipe-Soil Deformation Transfer Coefficient of Pipe Bend

According to the elastic foundation beam model [21], the pipe-soil deformation transfer coefficient  $\beta$  is:

$$\beta = \frac{1}{1 + \frac{EA}{K} \left( \frac{2\pi}{L} \right)^2} \quad (2)$$

where  $E$  is pipe's elastic modulus,  $A$  is pipe's cross-sectional area,  $L$  is seismic wave length, and  $K$  is foundation coefficient, which is the reaction force of the foundation soil on the unit length of the pipe when the foundation soil and the pipe are relatively displaced. When equation (2) is used to calculate the value of  $\beta$ , it is first applicable to straight pipe. And then it is applicable to the case that the angle between pipe direction and shear wave direction is 45°. The difficulty is to determine the value of  $K$ . In general, the value of  $K$  should be determined by in-situ experimental measurement. According to the Japanese anti-seismic criteria for chemical equipment, the value of  $K$  can be expressed as:

$$K = 3G = \frac{3\gamma_s}{g} V_s^2 \quad (3)$$

where  $G$  is soil's shear modulus,  $\gamma_s$  is bulk density of soil,  $g$  is acceleration of gravity,  $V_s$  is shear wave velocity.

Takada [22] conducted an experimental study on the axial restoring force between the pipe and the soil, and found that the pipe-soil friction under dynamic loading is about 70% of that under static loading, which can be taken as 1 MPa, 5 MPa and 10 MPa, corresponding to soft, medium and hard sites respectively.

In the 1970s, Japanese scholars used the theory of beam on elastic foundation and simple harmonic wave to put forward the analytical model of pipe-soil interaction, namely the response displacement method. The displacement transfer coefficient was divided into axial  $\beta_1$  and transverse  $\beta_2$  directions, as shown in the equation (4).

$$\beta_1 = \frac{1}{1 + \left( \frac{2\pi}{\sqrt{2L}} \right)^2 \frac{EA}{K}}, \quad \beta_2 = \frac{1}{1 + \left( \frac{2\pi}{L} \right)^4 \frac{EI}{K}} \quad (4)$$

Based on the previous pipe-soil deformation transfer theory and the seismic response test results of buried elbow in Table 2, considering the influence of wall thickness, pipe diameter, buried depth, soil properties, angle and other factors on the seismic resistance of elbow, and according to dimensional analysis and equation (4), with the method of control variable fitting, a calculation formula for estimating the pipe-soil deformation transfer coefficient at the maximum strain of buried elbow was put forward, as shown in the equation (5). And its correlation index is 0.95. Compared with the existing transfer coefficient formula of straight pipe, the influence of buried depth, wall thickness and elbow angle was considered in equation (5). Only El-Centro seismic wave was considered in this formula, so on this basis, Kobe seismic wave (predominant frequency 2.5 Hz) and artificial wave (For the artificial wave synthesized, the return period was 50 years, while its duration was 40 s with maximum and minimum periods of 4 s and 0.01 s, respectively. The dominant frequency of Fourier spectral curve for this wave is 4 Hz,) were loaded for the test. The peak value of wave was 0.8 g . Compared with the calculated value of equation (5), the relative error of the deformation transfer coefficient for the elbow obtained by applying Kobe wave and artificial wave was less than 20%, which indicated that the influence of seismic wave frequency on the deformation transfer coefficient at the elbow was small.

$$\beta = \frac{1}{1 + 0.005 \frac{EDt}{K} \left( \frac{2\pi}{L} \right)^2 u^{4.01} \left( \frac{H}{L} \right)^{-0.64} \cdot 1.72^{1.005 - 1.78 \cos \theta}} \quad (5)$$

where  $\beta$  is the pipe-soil deformation transfer coefficient at the maximum strain of the buried elbow;  $K$  is the foundation coefficient, the soft, medium and hard sites can be taken as 1 MPa, 5 MPa and 10 MPa respectively; in this paper, the clay was taken as 5 MPa, and the sand was taken as 3 MPa;  $L$  is the seismic wave length/m, the soil characteristic period was taken as 0.3 s, and the shear wave velocity mentioned above was about 120 m/s, and two thirds of the actual shear wave velocity was taken in this paper [21]. Therefore, for the actual situation, the seismic wave length was taken as 24 m, through the similarity ratio transformation, taken as 2 m for the test;  $D$  is pipe's diameter/m;  $t$  is pipe's wall thickness/mm;  $H$  is pipe's buried depth/m;  $\theta$  is the elbow angle/ $^\circ$ ;  $u$  is the pipe-soil friction coefficient, taking 0.4 in this paper. The research showed that [23] the selection of pipe-soil interface friction coefficient is related to soil physical and mechanical properties, soil type and pipe wall roughness. When the friction coefficient of pipe-soil interface is equal to 0.4, the axial force, circumferential strain and

pipeline shear stress are the smallest, and the seismic effect of pipeline is the best. Xiao and Huang [24] studied the friction coefficient at the pipe-soil interface. During earthquake, when the pipe-soil contact material is polyethylene and the sand has 0~16% water content, the dynamic friction coefficient at the interface is in the range of 0.4~0.6. Similarly, for clay, when the water content is 10%~50%, the friction coefficient at the interface is in the range of 0.37~1.10. For the above pipeline dynamic test, the interface friction coefficient of sand and clay was taken as 0.4. Therefore, in the finite element calculation model built in this paper, the tangential friction coefficient of the interface was set to 0.4.

Equation (5) is applicable to the calculation of pipe-soil deformation transfer coefficient at the elbow of buried cast iron elbow. When the pipe diameter is 0.5~1.5 m, the buried depth is 1~4 m, the wall thickness is 10~40 mm, and the elbow angle is greater than 30° of the pipe, the equation (5) has strong applicability to estimate the pipe-soil deformation transfer coefficient. It was found that when the elbow angle  $\theta$  was 180° and the pipe-soil friction coefficient  $u$  was 0.4, the buried depth  $H$  was 0.3 m and the shear wave length  $L$  was 2 m, the equation (5) could be simplified as equation (6):

$$\beta = \frac{1}{1 + 2 \times \left( \frac{2\pi}{L} \right)^2 \frac{EDt}{K}} \quad (6)$$

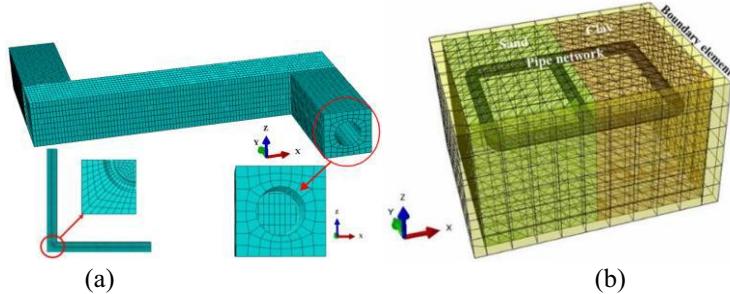
This equation is exactly the axial transfer coefficient  $\beta_1$  from equation (4), but the wall thickness  $t$  of the pipe is in m.

#### 4. Numerical Simulation Design Based on Goodman Contact Element

In the analysis of this paper, the three-dimensional model calculation results of ABAQUS finite element software were compared with the experimental calculation results. Taking the buried steel elbow as the research object, the pipe is API 5L X60 steel pipe. 210 GPa is for elastic modulus, 7800 kg/m<sup>3</sup> is for density, and 0.25 is for Poisson's ratio of the pipe. Ramberg Osgood's model was used to describe its nonlinear behavior. The specific size of the pipe was the same as that of the test converted by similarity ratio, and two groups of comparison models were added for each factor, as shown in Table 4. The physical parameters of the soil were the same as the test. In the simulation analysis, the Drucker-Prager model was used to describe the constitutive relationship of the soil, and the seismic wave input was also El-Centro wave.

Taking the bend diameter  $D = 1.0$  m as an example, in the numerical modeling, a bend with a radius of 0.5 m was drawn in the form of axis sweeping. S4R shell element was used for the elbow, 8 seeds were arranged along the circumferential direction of the pipe, and 2112 elements were divided into grids. At the pipe ends, a ring of viscoelastic boundary element was set to simulate the effect of rubber. In order to compare with the test, 19×14×12 m cube soil was cut and extended 1 m along the normal direction on its four sides and bottom surface. A uniform viscoelastic artificial boundary was established to simulate the infinite foundation. The soil around the pipe of 2 m×2 m centered on the pipe axis was selected for mesh refinement. Firstly, the soil around the pipe was

partitioned along the straight pipe section. Then, the soil around the pipe was partitioned along the elbow axis. Finally, the structured mesh division of the soil around the pipe was realized, with a total of 15360 elements. The soil around the pipe was also partitioned, and this part of a structured grid implemented was separated into 15604 elements by global seeding. Research showed that [25] the number of soil elements is about 30000, and the convergence of the element is the best. The mesh size of the artificial boundary was  $0.2 \times 0.2 \times 0.2$  m, with 584 elements. C3D8R solid element was used for soil and boundary, and the model grid was presented in Figure 6.



**Figure 6.** Finite element model for pipe bend buried. (a) Finite element mesh of soil around pipe. (b) Stereogram of finite element mesh.

Goodman contact element model [26]: Goodman proposed that the contact surface was completely consistent before the force was applied, and the thickness of the element at the contact surface was zero. For the ABAQUS software, only the frictional contact characteristics of the contact surface were considered, and the sliding deformation was transformed into two directions, namely parallel and vertical directions [27], as shown in Figure 7(a). If the interaction of shear stress in two directions is considered, the constitutive relationship of the contact surface is:

$$\begin{Bmatrix} \Delta\tau_1 \\ \Delta\tau_2 \end{Bmatrix} = \begin{bmatrix} k_{s1} & 0 \\ 0 & k_{s2} \end{bmatrix} \begin{Bmatrix} \Delta\gamma_1 \\ \Delta\gamma_2 \end{Bmatrix} \quad (7)$$

and  $k_{s1}$  and  $k_{s2}$  can be expressed as:

$$k_{s1} = \left( 1 - R_f \frac{\tau_1}{\sigma_n \tan \delta} \right)^2 K_1 \gamma_w \left( \frac{\sigma_n}{P_a} \right)^n \quad (8)$$

$$k_{s2} = \left( 1 - R_f \frac{\tau_2}{\sigma_n \tan \delta} \right)^2 K_2 \gamma_w \left( \frac{\sigma_n}{P_a} \right)^n \quad (9)$$

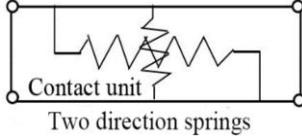
where  $K_1$ ,  $K_2$ ,  $n$ ,  $R_f$  are the non-linear index,  $\delta$  is the interface friction angle of the contact surface,  $\gamma_w$  is the volume density of water,  $P_a$  is the atmospheric pressure,  $\sigma_n$  is the normal contact stress,  $\Delta\tau_1$  and  $\Delta\tau_2$  are the shear stress increase in

two directions,  $k_{s1}$  and  $k_{s2}$  are the spring stiffness in both directions Direction,  $\Delta\gamma_1$  and  $\Delta\gamma_2$  are the strain increment in the two directions,  $\tau_1$  and  $\tau_2$  and are the shear stress in the two directions.

In Table 3, for specific parameter settings, the calculation method of each parameter was carried out by Wang et al. [28] in 2005.

**Table 3.** The Goodman contact parameters.

$K_1$	$K_2$	$n$	$R_f$	$\delta$ (°)	$\gamma_w$ (kN/m <sup>3</sup> )	$P_a$ (kPa)
2000	2000	0.56	0.74	29.3	10	100



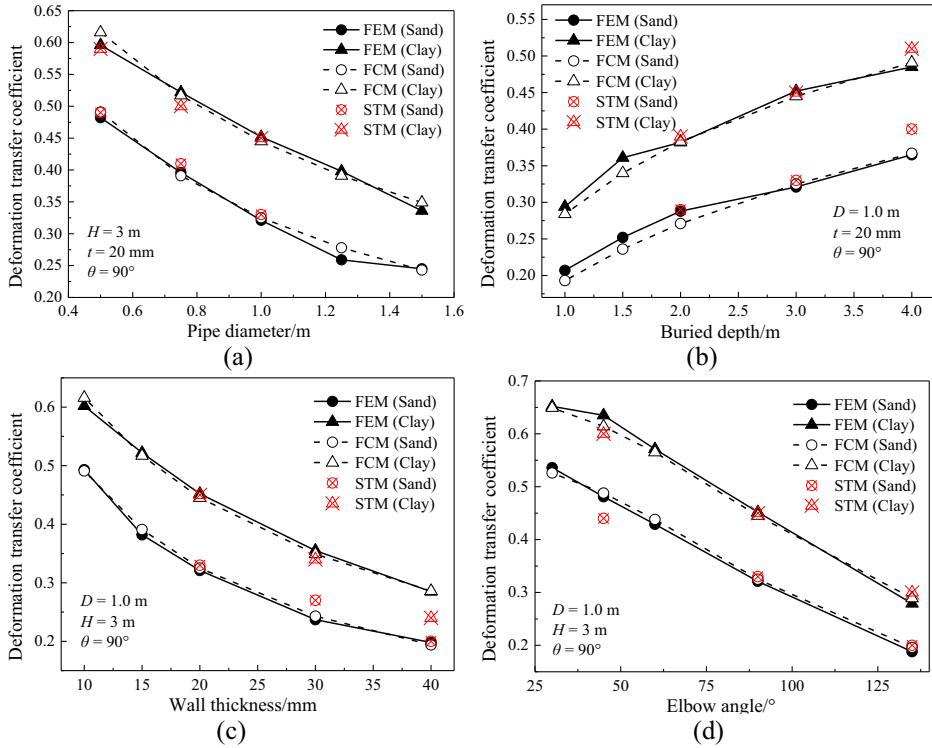
**Figure 7.** The Goodman contact element model.

## 5. Verification of the Pipe-Soil Deformation Transfer Coefficient Formula

In order to be consistent with the experimental measurement and analysis process, the displacement of the elbow was not extracted directly, but the acceleration of the elbow was extracted and integrated to calculate the displacement of the elbow. Statistics and calculations were made by the finite element method based on Goodman contact element and equation (5), and the pipe-soil deformation transfer coefficient was obtained, as shown in Table 4. The calculation results of the test were also added to the table, and the curves in Figure 8(a), (b), (c) and (d) were established.

**Table 4.** Statistics of the deformation transfer coefficient for pipe and soil obtained by two methods.

Pipe diameter /m	Buried depth /m	Wall thickness /mm	Angle /°	Finite element results		Formula calculation results	
				Sand	Clay	Sand	Clay
0.50	3.0	20.0	90	0.482	0.596	0.491	0.616
0.75	3.0	20.0	90	0.396	0.522	0.391	0.517
1.00	3.0	20.0	90	0.321	0.452	0.325	0.445
1.25	3.0	20.0	90	0.259	0.398	0.278	0.391
1.50	3.0	20.0	90	0.245	0.336	0.243	0.349
1.00	1.0	20.0	90	0.207	0.294	0.193	0.284
1.00	1.5	20.0	90	0.252	0.361	0.236	0.340
1.00	2.0	20.0	90	0.288	0.382	0.271	0.383
1.00	3.0	20.0	90	0.321	0.452	0.325	0.445
1.00	4.0	20.0	90	0.365	0.485	0.367	0.491
1.00	3.0	10.0	90	0.493	0.602	0.491	0.616
1.00	3.0	15.0	90	0.382	0.522	0.391	0.517
1.00	3.0	20.0	90	0.321	0.452	0.325	0.445
1.00	3.0	30.0	90	0.237	0.355	0.243	0.349
1.00	3.0	40.0	90	0.198	0.285	0.194	0.286
1.00	3.0	20.0	30	0.536	0.652	0.526	0.649
1.00	3.0	20.0	45	0.481	0.635	0.488	0.614
1.00	3.0	20.0	60	0.429	0.571	0.438	0.565
1.00	3.0	20.0	90	0.321	0.452	0.325	0.445
1.00	3.0	20.0	135	0.188	0.279	0.196	0.289



**Figure 8.** Comparison of finite element simulation (FEM), formula calculation (FCM) and scale test results (STM). (a) Pipe diameter  $D$ . (b) Buried depth  $H$ . (c) Wall thickness  $t$ . (d) Elbow angle  $\theta$ .

It can be found from Figure 8(a) that the pipe-soil deformation transfer coefficient decreases with the increase of pipe diameter under the condition of different pipe diameters, that is, the corresponding deformation of the elbow decreases when the soil is in large deformation, and the ability of the pipe to resist the deformation of the surrounding soil is enhanced. Therefore, the pipe diameter should be appropriately increased in the seismic design of the elbow. It can be seen from Figure 8(b) that the pipe-soil deformation transfer coefficient decreases with the increase of the buried depth of the elbow, that is, the corresponding deformation of the pipe increases when the soil has large deformation, and the ability of the pipe to resist the deformation of the surrounding soil decreases. Therefore, pipe bend should be buried shallowly in the seismic design of the elbow. As can be seen from Figure 8(c), the pipe-soil deformation transfer coefficient decreases with the increase of the wall thickness of the elbow, that is, the corresponding deformation of the pipe decreases when the soil has large deformation. That is to say, when the soil has large deformation, the corresponding deformation of the pipe decreases, and the ability of the pipe to resist the deformation of the surrounding soil increases; Therefore, the wall thickness should be increased properly in the seismic design of elbow. It can be seen from Figure 8(d) that the pipe-soil deformation transfer coefficient decreases with the increase of elbow angle, that is, when the soil has large deformation, the corresponding deformation of the pipe decreases, and the ability of the pipe to resist the deformation of the surrounding soil increases. Therefore, elbow with large elbow angle should be used as far as possible according to the actual situation of the project in the seismic design of elbow. The angle of the straight pipe is equivalent to

180°, so the seismic capacity of the straight pipe is the strongest. It can also be found from Figure 8 that the deformation transfer coefficient of the elbow in clay is larger than that in sand. The reason is that the foundation coefficient of clay is larger than that of sand. Therefore, the clay has a strong binding capacity to elbow during earthquake, making the elbow relatively dangerous. While the binding capacity of sand to the pipe is relatively weak, the pipe is relatively safe without sand liquefaction.

From table 4, it can be observed that the relative error of the pipe-soil deformation transfer coefficient calculated by the equation (5) and finite element simulation based on Goodman contact element is less than 5% under different working conditions. It indicates that the equation of pipe-soil deformation transfer coefficient proposed in this paper has certain accuracy for the different pipe diameters, buried depth, wall thickness, elbow angles and soil properties.

## 6. Conclusion

In this paper, shaking table scale test and Goodman contact element methods were used to study the pipe-soil deformation transfer coefficient under seismic and the change law of deformation transfer coefficient under multiple factors, the conclusions were as followed:

- The equation of deformation transfer coefficient of elbow fitting by shaking table test data has strong applicability, and when the elbow angle is 180 degrees, the equation can be simplified to the axial deformation transfer coefficient of existing straight pipe.
- The pipe-soil deformation transfer coefficient of elbow decreases with the increase of pipe diameter, wall thickness and elbow angle, and increases with the increase of buried depth. It shows that increasing pipe diameter, wall thickness and elbow angle and reducing buried depth are beneficial to the aseismic performance of elbow, and reducing soil hardness will also improve the aseismic performance of elbow.
- The relative error of the pipe-soil deformation transfer coefficient calculated by the equation and finite element simulation based on Goodman contact element is less than 5% under different pipe diameters, buried depth, wall thickness, elbow angles and soil properties, which means that the equation fitted of pipe-soil deformation transfer coefficient has certain accuracy.

## Acknowledgments

The supports from the National Key Technology R&D Program of China (grant number 2016YFC0802407) are greatly appreciated.

## References

- [1] Li J, Liu W. Large-scale urban network seismic reliability analysis and optimization. *Earthquake Engineering and Engineering Vibration*. 2006 Jun; 3(26):172-5.
- [2] Shiro T. Earthquake damage and countermeasures of lifeline system. *Special Structures*. 1997 Aug;14(4):25-30.
- [3] Guo ED, Yang D, Gao L, Liu Z, Hong GL. Study on practical method for earthquake damage prediction of buried pipeline. *World Earthquake Engineering*. 2012 Jun; 28(2):10-5.
- [4] Yigit A. Deprem etkisi altındaki gömülü sürekli boru hatları. İstanbul Teknik Üniversitesi Fen Bilimleri

Enstitüsü İstanbul Turkey. Doctoral Dissertation; 2015. p.15-9.

- [5] O'Rourke MJ, Liu X. Response of buried pipelines subjected to earthquake effect. Multidisciplinary Center for Earthquake Engineering Research, University at Buffalo Buffalo NY;1999.56 p.
- [6] Tang AP, Ou JP, Lu QN. Lifeline system network reliability calculation based on GIS and FTA. Journal of Harbin Institute of Technology. 2006 Dec; 13(4):398-403.
- [7] Hernandez-Fajardo I, Dueñas-Osorio L. Probabilistic study of cascading failures in complex interdependent lifeline systems. Reliability Engineering and System Safety. 2013 Mar; 111: 260–72.
- [8] Berezin Y, Bashan A, Danziger MM, Li DQ, Havlin S. Localized attacks on spatially embedded networks with dependencies. Scientific Reports. 2015 Mar;5:8934.
- [9] Wu BC, Tang AP, Wu J. Modeling cascading failures in interdependent infrastructures under terrorist attacks. Reliability Engineering and System Safety. 2016 Mar;147:1-8.
- [10] Ballantyne DB, Crouse CB. Reliability and restoration of water supply systems for fire suppression and drinking following earthquakes. National Institute of Standards and Technology (GCR Gaithersburg). 1997.195 p.
- [11] Romero N, O'Rourke TD, Nozick LK, Davis CA. Seismic hazards and water supply performance. Journal of Earthquake Engineering. 2010 Aug;14:1022–43.
- [12] Piratla KR, Ariaratnam ST. Criticality analysis of water distribution pipelines. Journal of Pipeline Systems Engineering and Practice. 2011 Feb; 2(3):91-101.
- [13] Trifunovic N. Pattern recognition for reliability assessment of water distribution networks. Ph.D. dissertation Delft Univ. of Technology (Delft Netherlands). 2012. 49 p.
- [14] Liu W, Xu L, Li J. Algorithms for seismic topology optimization of water distribution network. Science China Technological Sciences. 2012 Jul;55:3047-3056.
- [15] Liang JW, Xiao D, Zhang HW, Zhao XH. Optimal monitoring of pressure in water distribution system for health monitoring. Journal of Disaster Prevention and Mitigation. 2013 Dec;33(Suppl.):51-7.
- [16] Laucelli D, Giustolisi O. Vulnerability assessment of water distribution networks under seismic actions. Journal of Water Resources Planning and Management. 2014 Sep;141(6):04014082.
- [17] Ba ZN, Han YX, Liang JW. Risk assessment of the gas pipeline corrosion based on the improved AHP and fuzzy comprehensive evaluation method. Journal of Safety and Environment. 2018 Dec; 18(6):50-6.
- [18] Wang XJ, Guo ED, Yu TY, Li Q. Study on rapid evaluation method for seismic damage of gas pipeline network. World Earthquake Engineering. 2018 Mar;34(1):72-7.
- [19] Li Q, Gue ED, Li YQ, Liu ZB. Key problems of seismic resilience evaluation of water supply system. Journal of Catastrophology. 2019 Apr; 34(2): 83-8.
- [20] Ministry of Housing and Urban-Rural Construction of the People's Republic of China. Code for seismic design of buildings (GB 50011-2010). China Building Industry Press. 2016.36 p.
- [21] Ministry of Construction of the People's Republic of China. Code for seismic design of outdoor water supply, sewerage, gas and heating engineering (GB 50032-2003). China Architecture and Building Press. 2003. 54 p.
- [22] Takada S. Seismic response analysis of buried PVC and ductile iron pipelines. Proceedings of Pressure Vessel and Piping Conference (ASME). 1980 Jan: 23-32 .
- [23] Wang SJ, Zhu QJ, Liu YL, Wang HC. Research on earthquake resistance of buried pipeline under pipe-soil interaction. World Earthquake Engineering. 2007 Mar;23(1):49-52.
- [24] Xiao Y, Huang XY. Friction coefficient between buried hot oil pipeline and soil around pipe. Special Structures. 1990 Apr;7(4):43-6
- [25] Ryu DM, Lee CS, Choi KH, Koo BY, Song JK, Kim MH, Li JM. Lab-scale impact test to investigate the pipe-soil interaction and comparative study to evaluate structural responses. International Journal of Naval Architecture and Ocean Engineering. 2015 Jul; 7(4):720–38.
- [26] Kang F, Peng J. Detailed explanation of example for ABAQUS about geotechnical engineering. Posts and Telecommunications Press (Beijing).2017.138-42 p.
- [27] Manzari MT, Nour MA. On implicit integration of bounding surface plasticity models. Computers and Structures. 1997 May;63(3): 385-92.
- [28] Wang MS, Zhou XY, Hu YX. Studies on contact model of soil-pile dynamic interaction. Chinese Journal of Geotechnical Engineering. 2005 Jun; 27(6):616-20.

# Analysis on the Travel Characteristics of Sharing Bikes Connected with Different Types of Public Transportation Stations

Xiaohua YU<sup>1</sup>, Xiaohui WANG, Yanna ZHAO

*Jinan Rail Transit Group Co., Jinan, Shandong 250100, China*

**Abstract.** In order to better solve the problem of unbalanced supply and demand of connected shared bikes, this paper takes shared bikes as the research object, analyzes the usage characteristics of connected bikes in different types of public transport stations, and puts forward a data-based feature extraction method of shared bikes. Firstly, the usage data of shared bikes were collected, and the starting and finishing points were decoded. The public transport stations were divided into five typical types according to the decoded longitude, latitude and surrounding land types. Secondly, the connectivity activity, connectivity distance and user loyalty are put forward as the characteristic indicators of bike-sharing travel. Finally, taking the bicycle data of Chaoyang District of Beijing as an example, the travel characteristic indexes of shared bikes are analyzed. The results show that, as the "last kilometer" travel connecting tool of public transport, the peak of the use of shared bikes connecting residential stations is 6:30 to 9:30, and that of other stations is 7:30 to 9:30. The connecting distance of shared bikes is generally less than 2km, but the connecting distance of office sites can reach 3km, and this site has the highest user loyalty.

**Keywords:** Sharing bikes, public transportation stations, travel characteristics

## 1. Introduction

Relying on the advantages of wide distribution, flexibility and convenience, shared bikes meet the needs of passengers to travel in short distances and connect to public transportation, providing a new idea to solve the "last kilometer" problem, but at the same time, there are also problems such as random parking, unreasonable distribution and low utilization rate. Therefore, the analysis of the travel characteristics of shared bikes not only helps operators to understand the use characteristics of the bikes, but also lays a foundation for accurate demand prediction and scheduling of shared bikes in the later stage, which is conducive to improving the utilization rate of shared bikes.

In the existing studies, some scholars study the use data of shared bikes and analyze the travel characteristics of shared bikes connected to urban tracks and their role in the connection of urban tracks. Some foreign researchers [1-3] and others studied the factors affecting the use of shared bikes mainly from the aspects of users, environment, technology and so on, and the results revealed the great potential of developing the combined travel of rail transit and shared bikes. Zhou Rong, Yin Qiuyi,

---

<sup>1</sup>Corresponding author: Xiaohua Yu, Jinan Rail Transit Group Co., Jinan, Shandong 250100, China; E-mail: 18532327@qq.com.

Bai Yunyun et al. [4-6] analyzed the distribution, travel peak periods and cycling length of shared bikes around the subway. The research on cycling length was mostly based on questionnaire survey. Cheng Yuan [7] conducted a previous study and analyzed at the level of which station, indicating that shared bikes have a greater impact on residential and commercial areas. Some scholars focus on the study of models, including the demand prediction model of connected shared bikes and the scheduling model of shared bikes [8]. Qunchen [9] consider the bike rental behavior of each user, analyze the user's bike rental behavior, construct the dynamic evolution process of bike borrowing and return, and establish a mathematical model to determine the configuration of the bike fleet at each station when the bike is repositioned in the bike-sharing system.

Most of the existing studies focus on the travel characteristics of shared bikes, and there are few studies on the connection of shared bikes. The use characteristics of connected shared bikes are closely related to the surrounding properties and station types, but the research on this aspect is still insufficient. This study combines the characteristics of shared bikes with the types of public transport stations, making the configuration of shared bikes more refined, which is conducive to the maintenance of the sites in the later stage and of great significance for improving the utilization rate of shared bikes at public transport stations.

## 2. Data Preparation and Analysis Method

### 2.1. Basic Data Preparation and Processing

Bike-sharing began to rise until 2016, when it entered a stage of rapid development. 2017 became a significant year with the fastest growth rate of its user scale. Due to the increasingly prominent operation and management problems in the later period, the operating environment has changed, and the use of shared bikes has gradually gone to "decline". 2017 is the year of the fastest development of bike-sharing. The use data of bike-sharing is made public on the Internet, which can highlight the travel characteristics of most users, so it is very representative. Through data screening and statistics, basic data of shared bikes are obtained, such as number of stations, number of bikes. In order to study the travel characteristics of bike-sharing connected to different types of public transport stations, the basic data of this study adopted the bike-riding data of Chaoyang District, Beijing from May 10, 2017 to May 16, 2017 (Table 1). Due to the high site coverage rate, this paper takes several representative sites with high loan and return volume in Chaoyang District as examples to analyze and classify their land use properties. Firstly, the data shall be screened initially according to the requirements:

- (1) The data with obvious errors (such as data missing, incomplete, etc.) shall be deleted.
- (2) The lock threshold of bike-sharing switch is 1 minute. The data with the same USERID and two consecutive Start Time intervals of less than 1 minute are combined as one data.

**Table 1.** Data processing of Mobike.

Order ID	User ID	Bike ID	Bike Type	Start Time	Borrow Point Location	Return Point Location
163020	378959	22541	1	5/10 0:00	wx4f5pf	wx4f5pe
3428602	140756	354271	1	5/10 0:00	wx4g20b	wx4erbf
4580017	693933	459074	1	5/10 0:00	wx4ersk	wx4ersf
4117589	482046	420604	1	5/10 0:00	wx4dnv2	wx4dphv
2851526	895961	300008	2	5/10 0:00	wx4ff3z	wx4ff90

After preliminary data screening, the Geohash of the existing data is decoded into latitude and longitude by Python programming, and the result is shown in Table 2.

**Table 2.** Geohash decoded partial data.

Order ID	User ID	Bike ID	Borrow Point	Return Point	Borrow the bike to latitude and longitude	Return the bike to latitude and longitude		
163020	378959	22541	wx4f5pf	wx4f5pe	39.769821 17	116.50245 67	39.76844 788	116.5038 3
3428602	140756	354271	wx4g20b	wx4erbf	39.951095 58	116.36787 41	39.95109 558	116.3596 344
4580017	693933	459074	wx4ersk	wx4ersf	39.970321 66	116.35139 47	39.97306 824	116.3486 481
4117589	482046	420604	wx4dnv2	wx4dphv	39.756088 26	116.31294 25	39.75334 167	116.3307 953
2851526	895961	300008	wx4ff3z	wx4ff90	39.868698 12	116.47636 41	39.86457 825	116.4777 374

In order to ensure that the research data are within the boundary of Beijing, the geographic coordinates (39 "26" to 41 "03" north latitude, 115 "25" to 117 "30" east longitude) beyond the boundary of Beijing are deleted. For latitude and longitude, when accurate to 7 decimal places, the accuracy is 1cm. Therefore, latitude and longitude only need to be reserved to 7 decimal places.

## 2.2. Classification of Public Transport Stations Based on Land Use Types

There is an interactive relationship between transportation and land use. Different land uses around public transport stations have different degrees of attraction to passenger flow, and the use of surrounding connected shared bikes is also different. According to "urban land classification and planning construction land standards", land use is generally divided into commercial land, residential land, comprehensive land, industrial land and other land. According to the location of the public transport stations, the public transport stations are divided into commercial stations, residential stations, office stations, hub stations and school stations.

Python tools were used to output the public transport stations connected to shared bikes with more than 150 times per day, which are 13 stations, and the land use of their geographical locations were analyzed and classified according to the types of stations as shown in Table 3.

**Table 3.** Location and types of bike docking stations in Chaoyang District.

Serial Number	Connect to the rental	Longitude	Latitude	Public Transport	Geographical Position	Site type
1	wx4g4ch	116.4942169	39.9085235	Subway station	West exit of Beijing East	Hub
2	wx4g4cj	116.4955902	39.9085235	Subway station	East entrance of Beijing East	Hub
3	wx4g4e9	116.4791107	39.9222564	Bus station,	Beijing First Integrated Chinese	Business
4	wx4g53h	116.5161895	39.9085235	Bus station,	Quintessence	Office
5	wx4g559	116.5010833	39.9222564	Subway station	Garden Art Trading Suning Life Square	Business
6	wx4g55f	116.5024566	39.9236297	Bus station,	Near Building 33, Balizhuang Dongli	Residence
7	wx4g57t	116.5175628	39.9222564	Bus station,	Near Building 33, Gandew Garden	Residence
8	wx4g57v	116.5175628	39.9236297	Bus station,	Near Building 12, Gandew Garden	Residence
9	wx4g57y	116.5189361	39.9236297	Bus station,	Near Building 2, Gandew Garden	Residence
10	wx4g59n	116.5299224	39.9085235	Bus station,	Near Block A, No.5, Tangjia	Residence
11	wx4gh30	116.5546417	39.9085235	Bus station,	Jinxin Business Building	Office
12	wx4gh7y	116.5628814	39.9236297	Bus station,	Beijing Broadcasting	School
13	wx4ghc2	116.5766143	39.9098968	Bus station,	Xiliuxiang Village, Beijing East	Residence

### 2.3. Riding Distance Analysis Based on AutoNavi API

In actual road conditions, due to network constraints, different modes of transportation and different travel distances are adopted between the same starting and finishing points. The AutoNavi API path planning technology can put the start and end points on the map, and obtain the true cycling distance according to the road conditions. Compared with the existing Pythagorean Theorem and spherical triangle method, the result is accurate.

The application of AutoNavi API technology to calculate the cycling distance requires three steps: the first step is to apply for the "Web Service API" Key (Key); The second step is to concatenate the HTTP request URL and send the Key of the application together as the required parameter. The third step is to receive the data returned by the HTTP request (in either JSON or XML format) and parse the data.

## 3. The Characteristic of ConnectedShared Bike

### 3.1. Trip Characteristic Parameters

The travel characteristics of connected shared bikes should reflect the frequency and efficiency of the use of the bikes as well as users' habit of using the bikes. Therefore, the connectivity activity, connectivity distance and user loyalty indexes are selected as the travel parameters of shared bikes.

Connection activity refers to the number of shared bikes borrowed and returned for connection near a public transport station in a unit time, reflecting the frequency of use of shared bikes. The more shared bikes are borrowed for connection, the more active the site will be.

The connecting distance refers to the shortest distance from the start to the end of the Gaode map specified by users when connecting with shared bikes. The offset route between the start and end points is ignored. The riding distance method based on the Gaode map API is adopted for calculation.

User loyalty refers to the proportion of users with the frequency  $>1$  connected by shared bikes on the same site in the total number of users on the site within a unit time. The higher the user loyalty, the more fixed the bike-sharing users are.

### 3.2. Analysis of Mobike Trip Characteristics in Beijing

#### 3.2.1. Connection Activity Analysis

According to different statistical periods, the connection activity can be divided into two types: daily connection activity and hourly connection activity.

##### (1) Daily connection activity

Daily connection activity refers to the number of shared bikes borrowed and returned at the same site on a daily basis within a week. After a week of data analysis, the results are shown in Figure 1.

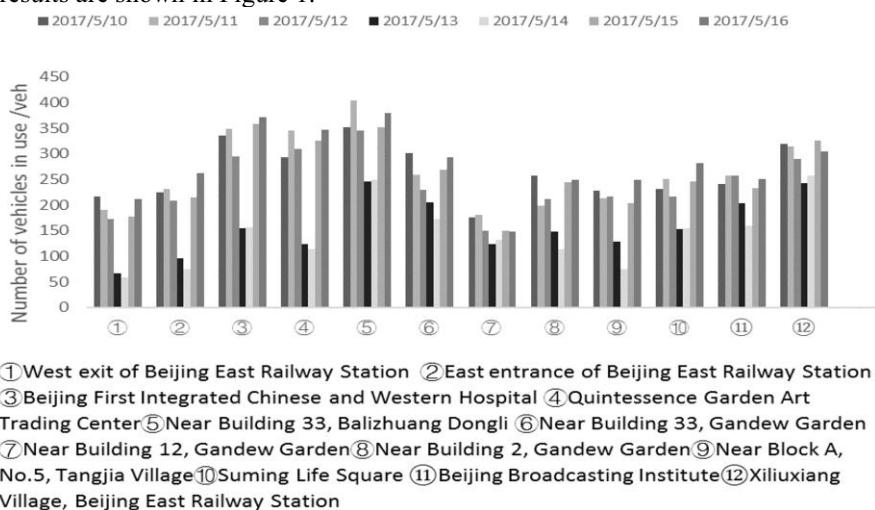


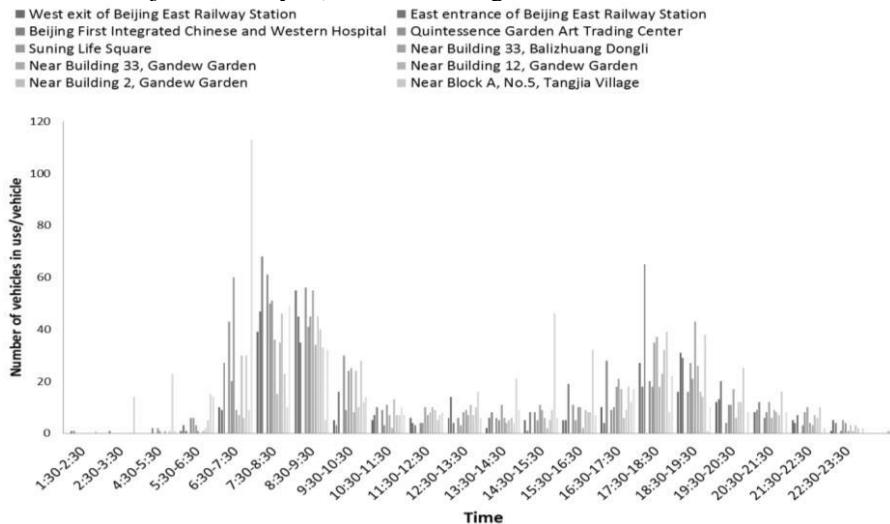
Figure 1. Daily connectivity activity of shared bikes at each site.

Figure 1 shows that in working days, the number of shared bikes in each site does not change significantly, which has little relationship with land use. However, the number of shared bikes in non-working days changes significantly compared with working days. The number of shared bikes used on non-weekdays was significantly lower than that on weekdays. There is no need to consider the time to travel outside work, and private cars are more comfortable to take public transportation, so most people will choose private cars to travel. Shared bikes are limited to the number of people who can take them at one time and the age of the users. Therefore, they are not

suitable for those who take a family as a unit on non-working days, and the number of times of use will be significantly reduced.

### (2) when the connection activity

Time connection activity refers to the number of shared bikes borrowed and returned at the site in an hour within a day. Select a day within a working week to draw the use of shuttle bikes at different types of stations within a day, and choose May 10 as the research object for analysis, as shown in Figure 2.



**Figure 2.** Connectivity of bike-sharing at each site.

For the 13 stations studied, there is a period of the highest connectivity activity in the morning and evening, which is basically consistent with the morning and evening peak hours, as shown in Figure 2. Due to the impact of housing prices in Beijing, most office workers live outside the city. Therefore, for residential sites, the connectivity activity ranges from 6:30 to 9:30 in the morning peak hours, while for other types of sites, the connectivity activity ranges from 7:30 to 9:30 in the morning peak hours. It shows that the morning rush hour of residential sites is earlier and lasts for a long time, which is consistent with the actual situation.

#### 3.2.2. Connection Distance Analysis

The connecting distance of shared bikes refers to the size of the influence domain of the connecting stations, namely the distance from the bike rental point of the starting place to the public transport station or from the public transport station to the destination bike rental point, reflecting the use efficiency of the bikes. The connection distance distribution is shown in Figure 3. Among all types of stations, the connection distance of more than 81.4% of single users is short distance travel, generally within 2000m. Due to the limitation of working time, about 17% of users in office stations ride a long distance, up to 3000m. For hub stations, 8% will choose to ride a longer distance, reaching 3500m. The number of cyclists over 4000m is almost zero.

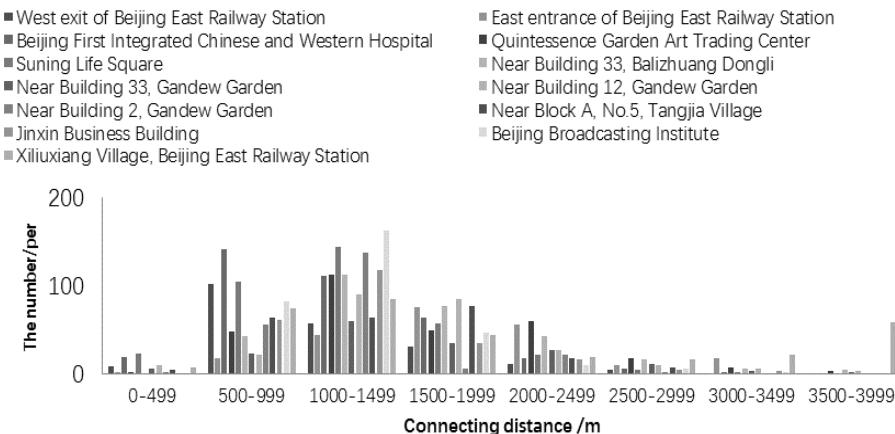


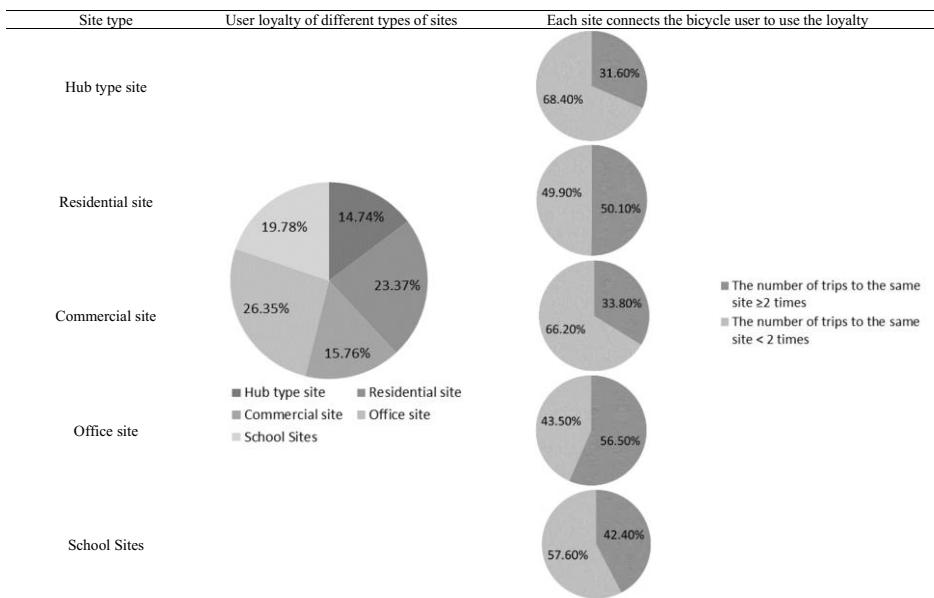
Figure 3. Schematic diagram of the distance between different types of stations.

### 3.2.3. User Loyalty Index

User loyalty reflects users' habit of using bicycles. In the data of the use of a bike, the USERID represents the unique code of the user, and the ID code of each user is fixed. The more times the same user connects to the shared bike at the same site, the higher the user's loyalty to the shared bike. From May 10 to May 16, 2017, the user loyalty of each type of site was analyzed, and the results were shown in Table 4.

In Table 4, office sites account for the highest proportion, followed by residential sites. Because most users of office sites are office workers, the time before work is tight. The speed of shared bikes is faster than that of walking, and there is no congestion in the process of driving. Therefore, most office workers are loyal to connect public transportation to office areas with shared bikes, and users of office sites have the highest loyalty. The main service object of residential site is residents, and the service group is limited, so the user loyalty is high. The school site is similar to the residential site in that its main service objects are students. However, because of the influence of schoolwork, students' spontaneity of travel is lower than that of residents, and their user loyalty is lower than that of residential sites. Commercial sites and hub sites serve the public, with large traffic and the lowest user loyalty.

Within a week, more than 50% of the people in office sites and residential sites use the same site more than one time, and nearly 50% of the people in school sites travel more than one time, indicating relatively high user loyalty. Hub sites have the lowest user loyalty, about 31.6%. Each site service group is different, has the great influence to the user loyalty. User loyalty requirements for the minimum amount of shared bikes should meet the needs of loyal users, and put forward requirements for the amount of shared bikes. At the same time, user loyalty can become an important index for operators to evaluate the fixed income of the site.

**Table 4.** User loyalty of different types of sites.

### 3.3. Suggest

Better connection planning based on the use characteristics of shared bikes can increase the number of shared bikes and promote the development of shared bikes. Based on the above analysis, the following suggestions are put forward:

(1) The analysis of connection activity shows that the morning rush hour of residential stations starts earlier and lasts longer than other stations. Therefore, operators can prioritize the scheduling of residential stations in the morning rush hour and reasonably arrange the priority of the scheduling order of shared bikes, so as to make the configuration of shared bikes more detailed.

(2) The cycling distance analysis of the shuttle bike shows that users mostly travel within 2km for a short distance. Therefore, the government should improve the network of public transportation lines, increase the coverage rate of bus stops, strengthen the publicity of "B+R", and make shared bikes play a bigger role, so as to relieve traffic congestion.

(3) The analysis results of user loyalty characteristics show that user loyalty is related to the stability of surrounding population. Residential and office sites have the highest population stability, so user loyalty is high. Operators should improve the comfort and cheapness of shared bikes, increase travelers' willingness to choose shared

bikes, and enhance users' loyalty to shared bikes, so as to make more travelers choose shared bikes.

#### 4. Conclusion

Shared bikes have been accepted by the public with the advantages of "on-demand, on-demand and on-demand", and play an important role in connecting public transportation. This paper analyzes the decoded data of shared bikes from three perspectives, namely, the connection activity, the connection distance and the user loyalty. It concludes that (1) there are still morning peak and evening peak when connected to shared bikes during working days, and the morning peak is more obvious;(2) Commuters are more inclined to ride for medium and long distances;(3) Different types of sites have different user loyalty, which is not only helpful for operators to understand the usage characteristics of bicycle users, but also of great significance to the research of scheduling time and scheduling quantity in the later period.

#### References

- [1] Sebastian M, Miguel RM, Francisco LC. User characteristics influencing use of a bicycle-sharing system integrated into an intermodal transport network in Spain. International Journal of Sustainable Transportation. 2020 Jul;14(7):513-24.
- [2] Gonzalez F, Melo-Riquelme C, Grange LD. A combined destination and route choice model for a bicycle sharing system, Transportation. 2016 May;43(3): 407-23.
- [3] Barna B, Zombor B, Péter I, Enikő ZV. Does Uber affect bicycle-sharing usage? Evidence from a natural experiment in Budapest. Transportation Research Part A. 2020 Jan; 133:290-320.
- [4] Zhou R, Wang YQ, Zhu L, Ji SJ. Research on travel characteristics of shared bikes based on temporal and spatial data. Journal of Wuhan University of Technology. 2019 Feb;43(01):159-63.
- [5] Yin QY, Zhen F, Luo SSX, Yan X, Guo SP. Influencing factors and layout suggestions of public bicycle travel space in new city. Urban Research. 2018 Dec; 12:9-15+46.
- [6] Chen QF, Fu ZY. Research on spatial-temporal characteristics of residents' cycling behavior under the influence of bike-sharing: A case study of central urban area of Nanjing. Urban Planning Society of China. 2018 Nov; 9:853-61.
- [7] Cheng Y, Xian K, Xu XY. Research on traffic distribution and users' cycling characteristics of shared bikes in Beijing. Institute of Urban Transportation. 2018 Oct; 10:810-19.
- [8] Nadav L, Chen G, Eran BE. An exploratory study of spatial patterns of cycling in Tel Aviv using passively generated bike-sharing data. Journal of Transport Geography. 2019 Apr; 76:325-34.
- [9] Chen Q, Liu M, Liu XY. Bike fleet allocation models for repositioning in bike-sharing systems. IEEE Intelligent Transportation Systems Magazine. 2018 Jan;10(1): 19-29.

# Study on Calculation Method of Contribution Rates of Water Pollution of Trans-Boundary Rivers

Xianzhong ZENG<sup>a</sup>, Shouning HAO<sup>b1</sup>

<sup>a</sup>Administrative Office of Songbaishan Reservoir, Guiyang 550025, Guizhou, China

<sup>b</sup>Xizang Agriculture and Animal Husbandry College, Linzhi 860000, Xizang, China

**Abstract.** The calculation of the contribution rates of the upstream and downstream administrative regions to water pollution of trans-boundary rivers is an important research subject. In this paper, a method based on the water quality mathematical model was proposed, which includes three steps: (1) Establishment of a model for a trans-boundary river; (2) Analysis of the pollution impact on the water quality at the control sections of the trans-boundary river; and (3) Calculation of the contribution rates to water pollution of the trans-boundary river. Taking Zhaosutai River, a trans-boundary river flowing through Jilin and Liaoning Provinces, as the study case, the contribution rates of these two provinces to water pollution of the river were determined by this method. The results showed that for the COD concentrations in the Control Sections of Laoqujiadian, Qijiazi and Dasijiazi of the main stream in January 2017, the contribution rates of the upstream pollution loads were 96.2%, 81.0% and 70.5%, respectively. The method of calculating the contribution rates to water pollution of trans-boundary rivers based on the mathematical water quality model proposed in this paper quantitatively can be used to identify the responsibilities of upstream and downstream administrative regions for water pollution of trans-boundary rivers.

**Keywords:** trans-boundary river; water pollution; liability identification; water quality model

## 1. Introduction

With the rapid development of society and economy, river environment is faced with increasing risks day by day [1, 2]. Since a trans-boundary river may involve two or more administrative regions, the study on its water environment improvement is particularly complicated [3-7]. Due to the different functional orientation and management requirements of the administrative regions in the basin of a trans-boundary river, the conflicts between different parties on water resource protection and the of water environment improvement of the river are increasingly prominent [8-10]. Wherein, the responsibilities of upstream and downstream administrative regions for water pollution at the control sections of the trans-boundary river are often the focus of disputes [11]. When the water quality from upstream is too poor, it is often difficult to ensure that the water quality at the control sections of the downstream meets the standard, even if the

---

<sup>1</sup>Corresponding author: Shouning Hao, Xizang Agriculture and Animal Husbandry College, Linzhi 860000, Xizang, China; E-mail: 645427848@qq.com

pollution load into the river from the downstream administrative region is low. To quantitatively define the responsibilities of upstream and downstream administrative regions for water pollution at the control sections and thus develop a plan for water pollution prevention and control of upstream and downstream of the river is a key research topic to solve the water pollution disputes and improve the water environment of a trans-boundary river. Moreover, more and more attention has been paid to the joint prevention and control of trans-boundary rivers in recent years. The explorations of joint pollution prevention and control, law enforcement and ecological compensation across the upstream and downstream administrative regions have been carried out in many river basins [10, 12, 13]. The calculation of the contribution rates of the upstream and downstream administrative regions to water pollution at the control sections is an important basis for these explorations. Therefore, it is of great significance to study the calculation method of the contribution rates of upstream and downstream administrative regions to water pollution at the control sections.

In previous research, the calculation of the contribution rates of upstream and downstream administrative regions to water pollution at the control sections was mostly based on the negotiations between the two parties and no quantitative calculation method available. These methods are arbitrary and lack of scientificity, which can not correctly identify the responsibilities of the upstream and downstream administrative regions, and can not provide scientific and feasible support means for the water pollution control of cross-border rivers. In recent decades, as a research method that can quantitatively simulate and predict the evolution process of water environment, the mathematical model method is more and more used in the field of water environment management. It is an important research direction to analyze the water pollution liability of transboundary rivers by the mathematical model method. In this paper, a calculation method of the contribution rates to water pollution of trans-boundary rivers based on the mathematical water quality model was proposed and verified with Zhaosutai River, a transboundary river flowing through Jilin and Liaoning Provinces as the study case. The research results of this paper provide a quantitative research means for calculation of the contribution rates to water pollution of trans-boundary rivers and a scientific base for the development of ecological compensation agreements and water pollution control of trans-boundary rivers.

## 2. Calculation Method of Contribution Rates

For trans-boundary rivers, the pollutant concentration in water at the control sections of the downstream is affected not only by the pollutant input from the downstream, but also by the input from the upstream. In the calculation of the contribution rates to water pollution in trans-boundary rivers, it is difficult and important as well to quantitatively identify the impacts of the pollutant inputs from upstream and downstream regions on the water quality at the control sections. The simulation based on the mathematical water quality model is an effective solution to solve this problem. Accordingly, the calculation method of the contribution rates to water pollution in trans-boundary rivers will be proposed in the following discussion, of which the schematic diagram is shown in Figure 1.

(1) Establishment of a hydrodynamic and water quality mathematical model for trans-boundary rivers. A hydrodynamic and water quality mathematical model for trans-boundary rivers is established by collecting the basic information of topography,

hydrology and water quality of the trans-boundary river and completing the calibration and verification of the model.

(2) Analyses of the impacts of the pollutant inputs from upstream and local region on the water quality at the control sections. Provided that the measured water quality at the control sections satisfies the principle of linear superposition, the mathematical relationship of the measured water pollutant concentration at the control section, the water pollutant concentration created by the input from upstream and that from the local administrative region is as follows:

$$C_{\text{mea}} = C_{\text{up}} + C_{\text{down}} \quad (1)$$

Where  $C_{\text{mea}}$  is the water pollutant concentration of a downstream control section, in mg/L;  $C_{\text{up}}$  is the water pollutant concentration measured at the upstream control section, in mg/L; and  $C_{\text{down}}$  is the water pollutant concentration created by the input from the local administrative region, in mg/L.

$C_{\text{up}}$  is calculated by simulation of the established hydrodynamic and water quality mathematical model with the following relevant conditions: the water pollutant concentration measured at the upstream control section is used as the upstream pollutant concentration, and the pollutant input from downstream is assumed to be 0. By subtracting  $C_{\text{up}}$  from  $C_{\text{mea}}$ , we can get  $C_{\text{down}}$ .

(3) Calculation of the contribution rates to water pollution of a trans-boundary river. For the water pollutant concentration at a downstream control section, the contribution rates of the upstream and downstream administrative regions are  $R_{\text{up}}$  and  $R_{\text{down}}$ , respectively. The responsibilities of the upstream and downstream regions for the water pollution of the trans-boundary river can be quantitatively defined, according to the following equations:

$$R_{\text{up}} = C_{\text{up}} / C_{\text{mea}} \quad (2)$$

$$R_{\text{down}} = C_{\text{down}} / C_{\text{mea}} \quad (3)$$

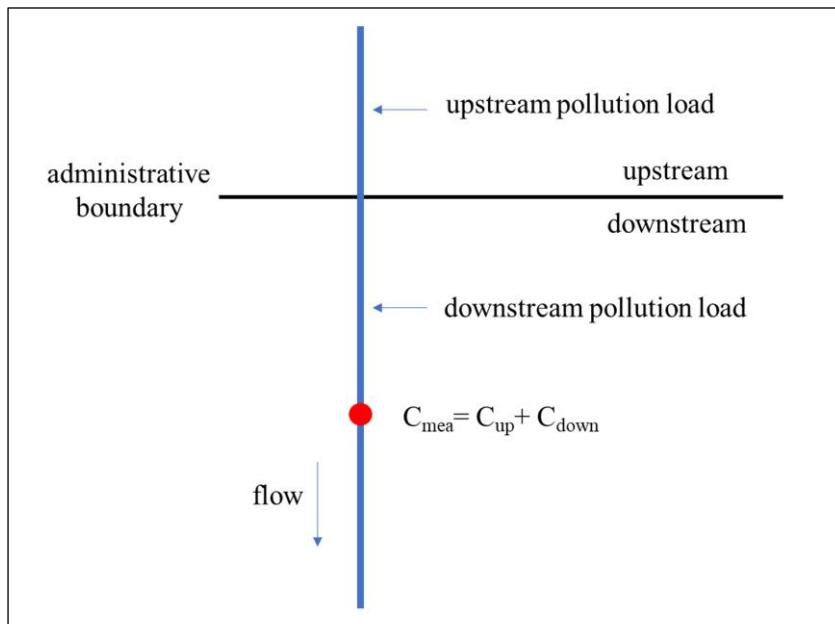
Where  $R_{\text{up}}$  is the contribution rate of the upstream administrative region to the pollutant concentration at the control section, and  $R_{\text{down}}$  is the contribution rate of the downstream administrative region.

### 3. Case Analysis

Taking Zhaosutai River, a trans-boundary river flowing through Jilin and Liaoning Provinces as the study case, this paper investigates the responsibilities of the two provinces for water pollution of this trans-boundary river by calculation of the contribution rates to the river water pollution with the hydrodynamic and mathematical water quality model mentioned above. The applicability of this model has been verified.

### 3.1. General Description of Zhaosutai River

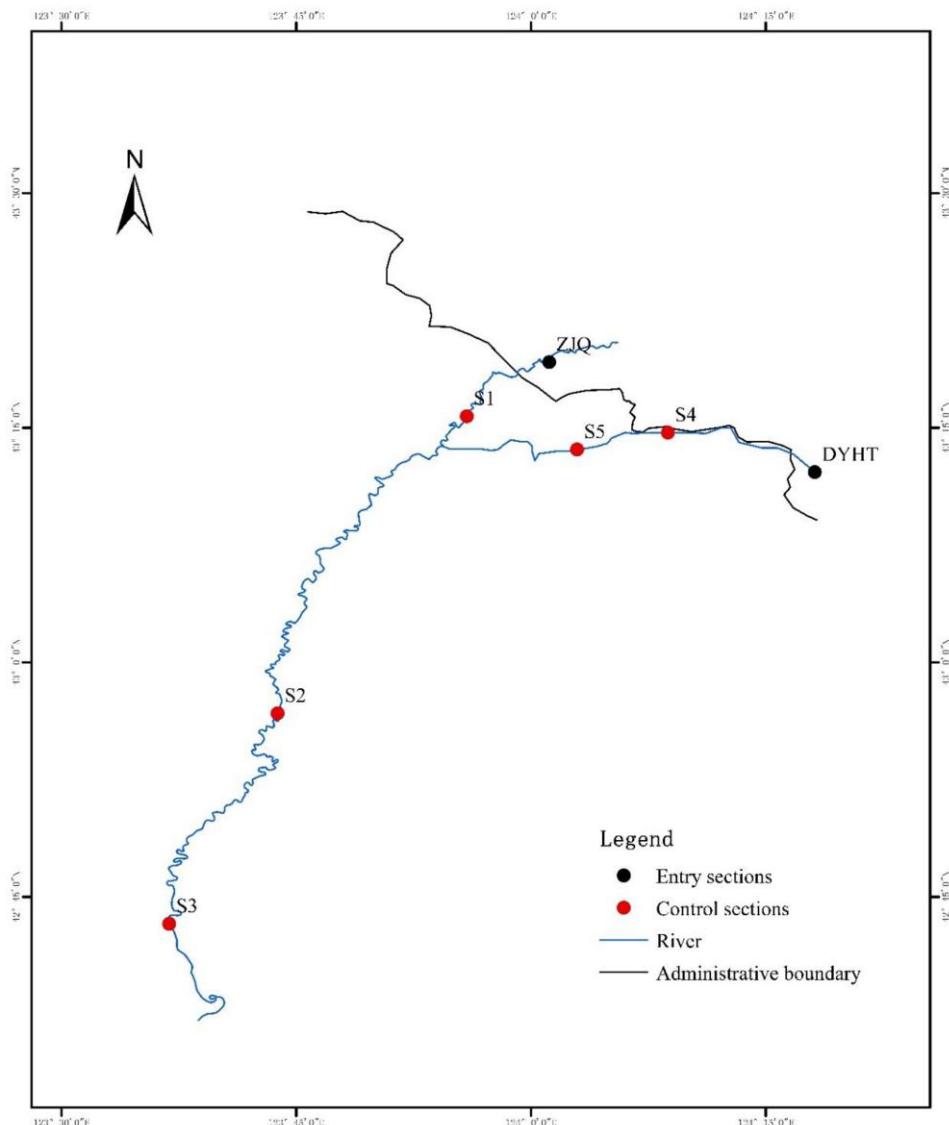
Zhaosutai River is the largest tributary of Liao River in Tieling region. It is adjacent to Dongliao River in the west, the watershed of Dongliao River in the north and east and Liangzi River in the south. The total length of main stream of Zhaosutai River is 213 km with a watershed area of 4583 km<sup>2</sup>. Zhaosutai River is in a mid-temperate zone with sub-humid monsoon continental climate. It has an average annual rainfall of 601.2mm, an average annual precipitation of 600mm, and an average annual temperature of 7.3°C.



**Figure 1.** Schematic diagram of calculation method for the contribution rates to the water pollution of a trans-boundary river.

Zhaosutai River is a trans-boundary river flowing through Jilin and Liaoning Provinces. It originates from the south of Heli Peak in the west of Dahei Mountain in Lishu County, Jilin Province and flows through Siping City, Jilin Province. The main stream of Zhaosutai River and its tributary Tiaozhi River flow through Changtu County, Tieling City, Liaoning Province and finally flows into Liao River in Tongjiangkou Town, Changtu County [14,15]. Zhaosutai River mainly flows through Changtu County of Liaoning Province. Its main stream in Changtu County is 157.82 km with a watershed area of 3004.1 km<sup>2</sup>. The schematic diagram of the water system of Zhaosutai River in Changtu County is shown in Figure 2.

There are five water functional zones along Zhaosutai River in Tieling City, including Jilin-Liaoning Riparian Buffer Zone, Heigang Agricultural Irrigation Zone, Huangjiuguan Agricultural Irrigation Zone, Tiaozhi River Jilin-Liaoning Riparian Buffer Zone and Tiaozhi River Linjia Agricultural Irrigation Zone. The control sections in the water functional zones are S1 (Laoqujiadian), S2 (Qijiazi), S3 (Dasijiazi), S4 (Linjia), S5 (Chenjiatun), respectively. The spatial locations of the control sections are given in Figure 2.



**Figure 2.** Schematic diagram of water system of Zhaosutai River in Changtu County.

### 3.2. Calculation of the Contribution Rates to Water Pollution of Zhaosutai River

#### 3.2.1 Establishment of the Hydrodynamic and Water Quality Mathematical Model for Zhaosutai River

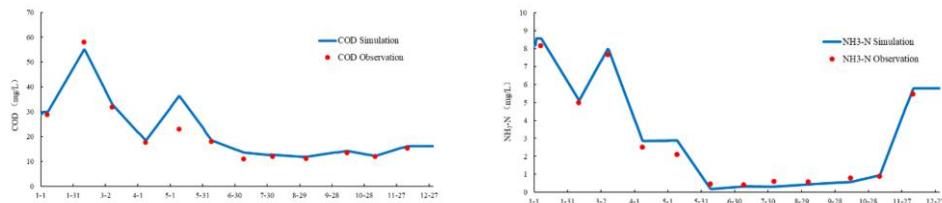
In this paper, the hydrodynamic and water quality mathematical model for Zhaosutai River was established based on HEC-RAS model. HEC-RAS is a hydrodynamic and water quality simulation software developed by the Hydrological Engineering Center of the U.S. Army Corps of Engineers, which is mainly used for the calculation of one-dimensional steady flow and unsteady flow in river courses and river networks. It has been widely used in the hydrodynamic and water quality simulation analysis of many rivers at home and abroad [16-20]. On the basis of HEC-RAS model, a hydrodynamic and water quality mathematical model of Zhaosutai River was established according to the water system situation and pollution sources of Zhaosutai River.

(1) Boundary Conditions and Initial Conditions. The upper boundary (i.e., the input point of upstream pollution load) conditions of the model were the daily flow rate and monthly water quality data of the provincial boundary control section, Zhangjiaqiao Section (ZJQ), where the main stream of Zhaosutai River enters Tieling City and those of the provincial boundary control section, Dongyihetun Section (DYHT), where Tiaozhi River enters Tieling City. The lower boundary condition of the model was the daily water level of the control section of Zhaosutai River where the river enters Liaoning Province. The input points of downstream pollution loads were the 6 generalized confluence points of the tributaries of Zhaosutai River and the sewage outfalls into the river in Tieling region. The initial condition of the hydrodynamic model was zero in flow velocity, and the initial condition of the water quality model was the measured water pollutant concentration.

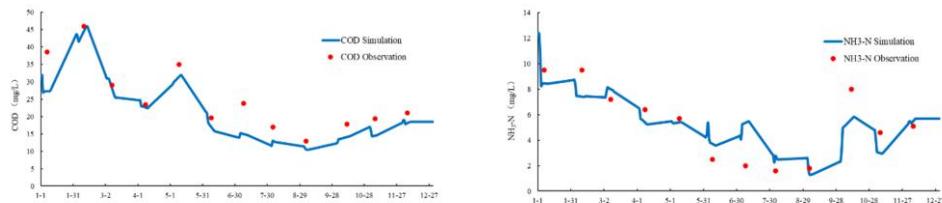
(2) Water Quality Indicators and the Model Step Size. COD and NH<sub>3</sub>-N were taken as the water quality indicators in the simulation of the mathematical model. According to the situation of data collection, January to December 2017 was chosen as the period for calibration and validation. Taking into consideration of both the calculation stability and efficiency, the calculation time step of the model was set to 5 minutes, and the length of the river course grid was set to 0.3-0.5km.

(3) Calibration and Validation of Model Parameters. The main parameters to be calibrated for hydrodynamic model and water quality mathematical model were river course roughness and the combined degradation coefficient of pollutants, respectively. In this paper, the roughness of natural river course was taken as the initial value of the roughness. By comparing the simulated and measured values of the water levels, the roughness of the main channel and the river beach are defined to be 0.027 and 0.018, respectively. The combined degradation coefficient was calibrated by comparing the measured water quality data at the control sections with the simulated water quality data. The combined degradation coefficients of COD and NH<sub>3</sub>-N were defined to be 0.17d<sup>-1</sup> and 0.13d<sup>-1</sup>, respectively.

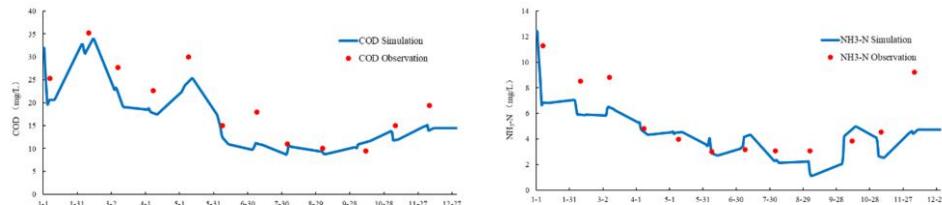
The control sections at Laoqujiadian, Qijiazi, Dasijiazi, Linjia and Chenjiatun were chosen as water quality calibration and validation locations. The comparison of simulated and measured values of COD and NH<sub>3</sub>-N for each location during the model calibration and validation period is given in Figures 3 to 7. It can be seen from the figures that the simulated values derived from the mathematical model agree well with the measured values of the water quality of Zhaosutai River. Therefore, the model can accurately simulate the distribution and transformation process of the pollutants in Zhaosutai River, and thus can be used to calculate the contribution rates from upstream and downstream to water pollution of Zhaosutai River.



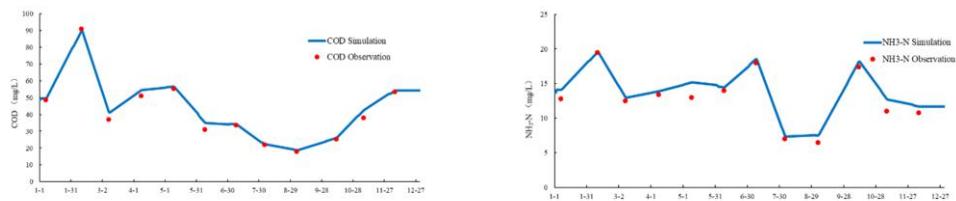
**Figure 3.** Comparison of simulated and measured values of COD and NH<sub>3</sub>-N at Laoqujiadian Control Section.



**Figure 4.** Comparison of simulated and measured values of COD and NH<sub>3</sub>-N at Qijiazi Control Section.



**Figure 5.** Comparison of simulated and measured values of COD and NH<sub>3</sub>-N at Dasijiazi Control Section.



**Figure 6.** Comparison of simulated and measured values of COD and NH<sub>3</sub>-N at Linjia Control Section.

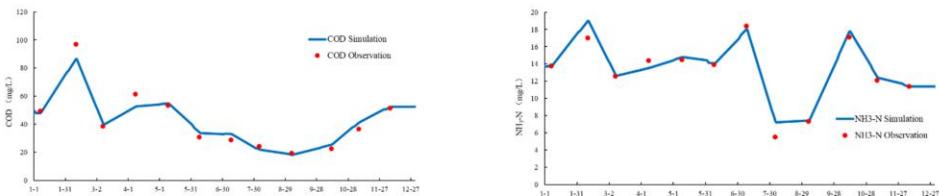


Figure 7. Comparison of simulated and measured values of COD and NH<sub>3</sub>-N at Chenjiatun Control Section.

### 3.2.2 Analysis of the Pollution Impact on Water Quality at the Control Sections of Zhaosutai River

The impact of the input of upstream and downstream pollution loads on the water pollutant concentrations of each of the control sections of Zhaosutai River in January of 2017 was analyzed through the established hydrodynamic and water quality mathematical model. The water pollutant concentration measured at the upstream control section in January of 2017 was taken as the input of upstream pollution load and assuming that the water pollutant concentration at the input point of downstream pollution load was 0, then  $C_{\text{mea}}$  was calculated by simulation under this condition. Provided that the water pollutant concentration at the input point of upstream pollution load was 0, and the measured concentration at the downstream control section in January 2017 was used as the input of downstream pollution load,  $C_{\text{down}}$  was calculated by simulation under this condition. The calculation results are given in Table 1 and Table 2.

It can be seen from Table 1 that for the main stream of Zhaosutai River, at Laoqujiadian Control Section in January of 2017 the COD concentration is 28.9mg/L, of which the COD concentration created by pollution load from upstream water is 27.8mg/L, and that created by the pollution load from downstream within Tieling region is only 1.1mg/L. It clearly shows that the pollution load from upstream water is the main factor affecting the water pollutant concentration at Laoqujiadian control section, while the downstream pollution load has little effect. The COD concentration at Qijiazi control section in January of 2017 is 38.6mg/L, of which the concentration from upstream water is 27.2mg/L, and the concentration from the downstream pollution load is 11.4mg/L, indicating that both upstream and downstream pollution loads have an important impact on the water quality at the control section, while the impact of upstream is greater than that of downstream. The COD concentration at Dasijiazi control section in January 2017 is 25.3mg/L, of which the concentrations from upstream and downstream pollution loads are 20.5mg/L and 4.8mg/L, respectively. The law of the impact of the upstream and downstream pollution loads on the pollutant concentration of this control section is similar to that of Qijiazi Control Section. For Tiaozi River, the COD concentration at the control section of Linjia in January 2017 is 13.6mg/L, of which the pollutant concentrations from the upstream and downstream pollution loads are 12.6 mg/L and 1.0mg/L, respectively. The COD concentration at Chenjiatun Control Section is 13.8mg/L, of which the pollutant concentrations from the upstream and downstream pollution loads are 12.1 mg/L and 1.7 mg/L, respectively. For the Control Sections of Linjia and Chenjiatun, the impact of the upstream pollution loads on the water quality in

the control sections is significantly greater than that of the downstream pollution loads. The impact of upstream and downstream pollution loads on NH<sub>3</sub>-N concentration of Zhaosutai River is similar to that on COD concentration.

**Table 1.** Analysis of impact on COD concentrations at the five control sections of Zhaosutai River in January 2017.

River	Control Section	$C_{\text{mea}}$ (mg/L)	$C_{\text{up}}$ (mg/L)	$C_{\text{down}}$ (mg/L)
Main Stream of Zhaosutai River	Laoqujiadian	28.9	27.8	1.1
	Qijiazi	38.6	27.2	11.4
	Dasijiazi	25.3	20.5	4.8
Tiaozi River	Linjia	48.7	46.5	2.2
	Chenjiatun	49.5	45.2	4.3

**Table 2.** Analysis of impact on NH<sub>3</sub>-N concentration at the five control sections of Zhaosutai River in January 2017.

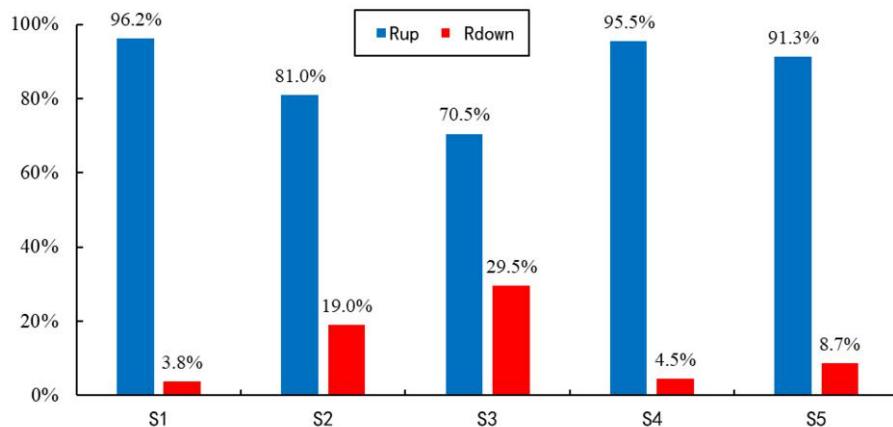
River	Control Section	$C_{\text{mea}}$ (mg/L)	$C_{\text{up}}$ (mg/L)	$C_{\text{down}}$ (mg/L)
Main Stream of Zhaosutai River	Laoqujiadian	8.2	8.0	0.2
	Qijiazi	13.5	8.5	5.0
	Dasijiazi	11.3	6.8	4.5
Tiaozi River	Linjia	13.6	12.6	1.0
	Chenjiatun	13.8	12.1	1.7

### 3.2.3 Calculation of the Contribution Rates to Water Pollution of Zhaosutai River.

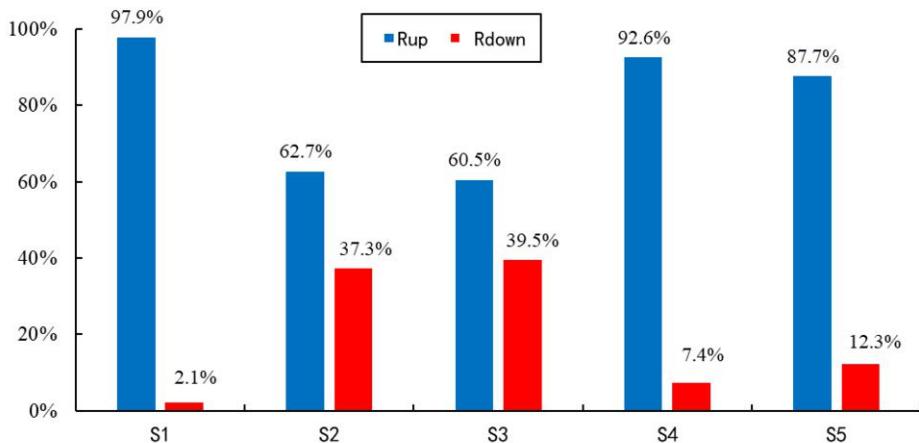
Based on the analysis results of the impact on the water quality at each of the control sections of Zhaosutai River, the responsibilities of upstream and downstream pollution loads for water pollution of Zhaosutai River is clearly defined, as shown in Figure 8 and Figure 9. It can be seen from Figure 8 that for the COD concentrations in the Control Sections of Laoqujiadian, Qijiazi and Dasijiazi of the main stream of Zhaosutai River in January 2017, the contribution rates of the upstream pollution loads are 96.2%, 81.0% and 70.5%, respectively, while those of the downstream pollution loads are 3.8%, 19.0% and 29.5%, respectively. For the COD concentrations in the control sections of Linjia and Chenjiatun of Tiaozi River, the contribution rates of the upstream pollution loads are 95.5% and 91.3%, respectively, while those of the downstream pollution loads are 4.5% and 8.7%, respectively. It can be seen from Figure 9 that for NH<sub>3</sub>-N concentrations at the Control Sections of Laoqujiadian, Qijiazi and Dasijiazi in the main stream of Zhaosutai River in January 2017, the contribution rates of the upstream pollution loads are 97.9%, 62.7% and 60.5%, respectively, while those of the downstream pollution loads are 2.1%, 37.3% and 39.5%, respectively. For the Control Sections of Linjia and Chenjiatun of Tiaozi River, the contribution rates of the upstream pollution loads are 92.6% and 87.7%, respectively, while those of the downstream pollution loads are 7.4% and 12.3%, respectively.

From the above analyses, it can be seen that the responsibilities of the upstream pollution loads for the pollutant concentrations in the control section are inversely

proportional to the distance between the control section and the input points of the upstream pollution loads. The sequence of the responsibilities of the upstream pollution loads of Zhaosutai River for the pollutant concentrations in each of the control sections is as follows: Laoqujadian > Qijiazi > Dasijiazi. Moreover, the responsibility of the upstream pollution load in each of the control sections of Zhaosutai River is greater than that of the downstream pollution load. This indicates that the key to water pollution control and water environment improvement of Zhaosutai River is to improve the upstream water quality.



**Figure 8.** Calculation Results of the contribution rates of the upstream and downstream administrative regions of Zhaosutai River to COD concentrations in January 2017.



**Figure 9.** Calculation results of the contribution rates of the upstream and downstream administrative regions of Zhaosutai River to  $\text{NH}_3\text{-N}$  concentrations in January 2017.

## 4. Conclusions

In this paper, a method for calculation of contribution rates to water pollution of trans-boundary rivers was proposed based on the water quality mathematical model. The applicability of this method was validated with data from Zhaosutai River, a trans-boundary river. The main conclusions of the research are as follows:

(1) The method of calculation of the contribution rates of water pollution of trans-boundary rivers based on the water quality mathematical model mainly includes three steps: (a) establishing a mathematical model of water quality for the trans-boundary river; (b) analysing the pollution impact of the upstream and downstream pollution loads on the water quality in the control sections based on the water quality mathematical model; and (c) defining the responsibilities of the upstream and downstream pollution loads for the pollutant concentrations in the control sections according to the analysis results.

(2) The responsibilities of the upstream pollution load for the pollutant concentrations in the control section are inversely proportional to the distance between the control section and the input points of the upstream pollution loads.

(3) The responsibilities of the upstream pollution load in all control sections of Zhaosutai River are greater than those of the downstream pollution loads. This indicates that the key to water pollution control and water environment improvement of Zhaosutai River is to improve the upstream water quality and reduce the input of upstream pollution loads.

Though a quantitative calculation method of the contribution rates to water pollution of trans-boundary rivers was proposed, further research is needed on application of this method to water resources protection and water environment improvement, including development of ecological compensation agreements for trans-boundary rivers and the comprehensive control of water pollution of trans-boundary rivers.

## References

- [1] Xue XUE, Linong WANG, Haoran XING, Yu ZHAO and Zaizhao WANG. Characteristics of phytoplankton-zooplankton communities and the roles in the transmission of antibiotic resistance genes under the pressure of river contamination. *Science of The Total Environment*. 2021 Aug;780: 146452.
- [2] Kang WANG, Pengxiang WANG, Renduo ZHANG and Zhongbing LIN. Determination of spatiotemporal characteristics of agricultural non-point source pollution of river basins using the dynamic time warping distance. *Journal of Hydrology*. 2020 Apr;583: 124303.
- [3] Qiting ZUO, Bingbing WU, Wei ZHANG and Junxia MA. A theoretical method of water diversion for trans-boundary rivers and calculation of new scheme of water diversion for Yellow River. *Resources Science*. 2020 Jan; 42(01): 37-45.
- [4] Lei YU, Libin LIU, Ruirui ZHU, Hailan LIN, Pei LIU, Rilong ZHU and Haiting PAN. Case Study on emergency monitoring of thallium pollution in trans-boundary rivers. *Environment and Development*. 2020 Sept;32(09): 138-9.
- [5] Yang LI, Yuchi LUO, Jian XIAO and Yejia LV. Study on environmental protection countermeasures for drinking water sources of trans-boundary rivers in Guangdong Province. *Environmental Science and Technology*. 2019 Oct; 32(05):58-62.
- [6] Ying TIAN, Beilei JIANG, Huojian HUANG and Fuxin SHEN. Study on risk factor identification and risk control of trans-boundary river safety. *Water Resources and Hydropower Engineering*. 2018 May;49(05): 9-15.
- [7] Yong ZHONG, Hui LIU, Fuqiang TIAN, and Mu LIN. Prisoner's dilemma in trans-boundary river cooperation and the approach of cooperation evolution. *Journal of Hydraulic Engineering*. 2016 May; 47(05): 685-92.
- [8] Jingjing DAI, Zhaoyi SHANG and Haoyang LI. Study on the ways of trans-boundary cooperation management for Taipu River. *Water Resources Protection*. 2016 Jan; 32(01): 142-7.

- [9] Fangzhou CHEN and Ruifang WANG. Study on long-term effect of ecological compensation mechanism for Xin'anjiang River Basin. *Yangtze River*. 2021 Feb;52(02): 44-9.
- [10] Jin CHEN and Zhengjie YIN. Scientific subjects and countermeasures of ecological compensation for Yangtze River Basin. *Journal of Yangtze River Scientific Research Institute*. 2021 Feb;38(02): 1-6.
- [11] Hong XING. Discussion on the settlement mechanism of trans-boundary water pollution disputes in Guangdong Province. *Journal of Guangxi University for Nationalities (Philosophy and Social Sciences Edition)*. 2013 Jan;35(01): 176-80.
- [12] Meichen PAN and Bo SONG. Willingness to accept compensation should be taken as the upper limit of ecological compensation standard. *China Environmental Science*. 2021 Feb;39(01):1-8.
- [13] Tiangu LV, Fangping LIU, Li WANG and Anni KONG. Identification, causes, mechanism and countermeasures of water resources management conflicts in trans-boundary watersheds: a case study of Poyang Lake watershed. *Yangtze River*. 2021 Feb;52(02): 85-89.
- [14] Xin RAN, Shijun SUN, Jiang FENG, Ming CUI, Zhenyu LONG and Sihan LIU. Study on performance optimization of water environment management based on control of total amount of pollutants: a case study of Zhaosutai River watershed in Jilin Province. *South-to-North Water Transfers and Water Science and Technology*. 2018 May;16(04): 128-135+145.
- [15] Yong ZHANG, Jie TANG, Zhaoyang LI, Yuan BAI, Yu PAN and Xiaojiao WANG. Study on characteristics of agricultural non-point source pollution in Zhaosutai River watershed in Jilin Province *Environmental Pollution & Control*. 2015 Dec;37(12): 29-34+40.
- [16] Bo WU, Xuzhao WANG, Hongfei ZANG and Jianyun WANG. Determination of inundation range of mountain torrents in villages along Xiaodongchuan River watershed based on HEC-RAS and GIS. *Water Resources and Power*. 2016 Sept; 34(09): 52-5.
- [17] Adeva-Bustos A, Hedger R D, Fjeldstad H P, Stickler M and Alfredsen K. Identification of salmon population bottlenecks from low flows in a hydro-regulated river. *Environmental Modelling & Software*. 2019 Oct;120: 104494.
- [18] Yongjia SONG and Daye WANG. Application and research of HEC-RAS model in water surface profile deduction for hills-and-plains compound river course in small watershed *China Rural Water and Hydropower*. 2020 Feb;(03):146-9.
- [19] Xingnan ZHANG and Shunfeng PENG. Research and application of flood routing simulation system in plain areas. *Journal of Hydraulic Engineering*. 2010 July ;41(07) :803-9.
- [20] Xiaogang LI, Cunchang HUANG and Yuzhu ZHANG. Study on flood hydrology of Wuding River at ten-thousand-year scale based on HEC-RAS model *Resources and Environment in the Yangtze Basin*. 2020 May;29(02): 526-34.

# The Impact of Air Pollution on the Development of Inbound Tourism in Chinese Megacities: The Case Study of Beijing, Guangzhou and Chongqing

Nianlin ZHOU<sup>1</sup>, Yeli GU and Manyuan JIANG

*South China Business College, Guangdong University of Foreign Studies, 181*

*Liangtian M. Rd., Baiyun District, 510545 Guangzhou, China*

**Abstract.** The existing studies pay more attention to the impact of public transport and other public service facilities on urban air pollution and tourism, but less on the negative effect of air pollution caused by carbon emissions of business fixed investment on inbound tourism. This article attempts to make a supplementary analysis about the above point through examining the correlation between air pollution associate with business fixed investment and the size of inbound tourism based on panel data of three megacities (Beijing, Guangzhou and Chongqing) in China over the period from 2015 to 2019. The findings of this paper show that the effects of air pollution linked with carbon emissions from business fixed investment on the number of inbound tourists (NIT) is a negative correlation, while the influence of GDP per capita and tourism revenue on NIT reveal a positive relationship by applying fixed effects model for benchmark regression and the system-GMM estimator for robustness check. Moreover, the negative influence of PM 10 on sample cities is more than PM2.5. Some different results of core variables between benchmark and sub-sample regressions don't imply the above conclusion to be substantively changed because of different distribution and concentration of nominal inbound tourists in specific sample megacities. In order to fundamentally improve air quality and to stimulate the development of inbound tourism, the suggestion of this study is to promote new business fixed investment with clean energy of renewable and low carbon.

**Keywords:** Air pollution, impact, inbound tourism, megacities

## 1. Introduction

In current global transition period in which pursues sustainable development, it has become a global agenda for countries to stimulate low-carbon economic growth and eco-tourism through promoting clean energy. Finnish government, for example, is developing new technologies and smart practices through offering diverse subsidies, incentives and distributing biogas and electric fuels for businesses and citizens, so that push forward disusing fossil fuels. Denmark generated half of its electricity from solar and wind power in 2019 and is trying to realize 100 % renewable energy in the future [1]. Sweden is promoting the target of zero emissions of greenhouse gases by 2045 by

---

<sup>1</sup>Corresponding author: Nianlin Zhou, South China Business College, Guangdong University of Foreign Studies, 181 Liangtian M. Rd., Baiyun District, 510545 Guangzhou, China; E-mail: gwjdyjzx@163.com.

introducing a climate policy framework as per the climate goals with a set of orderly and stable emission reduction actions [2]. Chinese government announced to reduce 60%-65% of carbon intensity by 2030 and to drive carbon emission to zero by 2060 [3]. The aims of decreasing carbon emission of different countries are engendering a profound influence on the model of economic growth and the development of tourism.

Megacities are the centers of enhancing business fixed investment for economic growth and expanding inbound tourism development in China. On the one hand, the improvement of tourism infrastructure and service facilities need to rely on business fixed investment. On the other hand, air pollution linked with high emissions by way of investment on fixed asset leads to a negative impact on the size of inbound tourism. How to lower air pollution to enlarge inbound tourism and to keep sustainable economic growth would be the focus of research.

This article examined the correlation between the size of inbound tourists and air pollution associated with carbon emission from business fixed investment based on the empirical analysis of Chinese megacities, Beijing, Guangzhou and Chongqing. The purpose is to suggest that one of fundamental ways to cut down air pollution and to expand the size of inbound tourist must proceed from promoting technological innovation of new energy and new infrastructure investment.

The possible contribution of this study is to introduce investment on fixed asset linked with carbon emissions into the research that focuses on improving air pollution and promoting inbound tourism development. The existing studies pay more attention to the impact of public transport and other public service facilities on urban air pollution, but less on the negative effect of air pollution caused by carbon emissions of business fixed investment on inbound tourism. This paper attempts to make a supplementary analysis.

## **2. Literature review**

As the inbound tourism centers, megacities have been widely studied about their functions of providing services for tourists, environment condition of developing tourism, and the influence of air quality on tourists.

Some scholars stress the functions and contribution of urban tourism. Wan et al. [4] point out that tourist megacities have comprehensive functions to improve capacities of tourism infrastructure, services, destination management and ecological environment condition through policy coordination on the one hand, and to provide multifunction services in tourist destinations, culture entertainment, restaurants, public facilities for tourists on the other [5-6]. The boom of urban tourism has also made contribution to increase employment, consumption and local economic growth [7]. Others describe a whole picture about the elements affecting urban tourism, in which include rank and abundance in tourism resources, service facilities for tourists, tourism infrastructure, the quality of tourist products and information about tourism destination influenced on tourist's decisions or choices [5].

Most of all, studies have paid common attention to the significance of environmental and air state that may influence on tourist choice and urban tourism development [5,8-9]. Since the development of tourism heavily depends on high-quality ecological environment and natural resources [10], air quality has been one of the important indicators to measure whether a city reaches the standard of international tourism city [11]. Moreover, arguments discuss different types caused

environmental pressure in tourist cities. Tourist traffic, for example, with increasing travel of tourists from one city to other, has led to high exhaust emissions and environmental problems via energy consumption of public transportation [12-15]. In particular, the main categories of air pollution, including dusts (PM2.5 and PM10), tropospheric ozone, sulfur oxides (SO) and nitrogen oxides, which definitely impact on urban tourism, are described in academic circles [16-18]. The monthly average concentrations of SO<sub>2</sub>, PM2.5 and PM10 have a more significant negative externality on the scale of inbound tourism in megacities than in non-central cities [19]. Moreover, the concern of potential overseas tourists about air quality and haze goes beyond the attraction of core attractions, such as history, culture and nature, and becomes the main factors hindering their intention to come over to China [20]. The risk perception of tourists, that is, fear and anxiety about the potential risks of a tourist destination, would drive them to give up visiting a certain place [21]. Saenz concludes that the increase and decrease of tourists' daily stock is an important indicator to evaluate the level of air pollution in urban [22].

In terms of sources of air pollution, literatures have studied the impact of public transportation on inbound tourism, but a few have analyzed the influence of business fixed investment on tourism. One more important function of Chinese megacities is to promote inbound tourism and economic growth by increasing investment on fixed assets. In other words, the issue of air pollution in megacities mirrors a common situation of relying on fossil fuels for business fixed investment heavily at the present stage for stimulating both economic growth and inbound tourism in China [23]. A number of empirical analysis indicate that the aggravation of air pollution is associated with increasing investment on fixed assets (IFA), power consumption and vehicle exhaust emission due to industrialization and urbanization of scale expansion in China [24-26]. Particularly, among six kinds of greenhouse gases in industrial carbon emissions, there are five of which are related to PM2.5 [27].

Hence, the impact from both air pollution and business fixed investment with carbon emission on the size of inbound tourists would be taken into account in this study.

### 3. Methods

#### 3.1. Data and Variables

##### 3.1.1 Data Sources

This paper uses a sample consisting of a panel data with 180 observations in three major tourist cities, Beijing, Guangzhou and Chongqing, in China over the period from 2015 to 2019 on monthly based. The data of the numbers of inbound tourist (NIT) are obtained from the published official database of bureaus of culture, radio, television and tourism in Beijing, Guangzhou and Chongqing separately. We collected the air monthly data of urban in which include the air pollution of particulate matter (PM2.5), inhalable particulate matter (PM10) and sulfur dioxide (SD) in the sample cities according to the published data from environmental monitoring of China. The data of GDP per capita (AGDP), investment on fixed assets (IFA), tourism revenue (TR) and consumer price index (CPI) are gotten from bureaus of statistics of these cities, and foreign exchange rate between RMB and US dollar (FER) from state administration of foreign exchange of China. We used STATA15 program for the empirical analysis.

### 3.1.2 Variables

*Dependent Variable: The Number of Inbound Tourists (NIT).* We take the number of inbound tourists as the dependent variable because it is a widely accepted indicator to explain the scale of inbound tourism [28] and the extent of sustainable development of tourism in international tourist cities [29].

*Independent variable: total air pollution (TAP).* It includes particulate matter 2.5(PM2.5), inhalable particulate matter 10 (PM10) and sulfur dioxide (SD). Air pollution is recognized as the main downturn cause of inbound tourism in China [30]. The monthly data of air pollution from 2015 to 2019 are selected as the core explanatory variables for the sample cities.

*Control variables: GDP per capita (AGDP), investment on fixed assets (IFA), tourism revenue (TR), consumer price index (CPI) and foreign exchange rate (FER).* We introduce a number of control variables that are known by both Chinese and foreign literature for influencing on tourism development and the number of inbound tourists in megacities. For instance, GDP per capita is used as a characterization index of economic development level, which is obtained by GDP dividing by resident population of sample megacities. In addition, investment on fixed assets (IFA) is an important control variable. Because there are four major industrial sectors, that is, energy, industry, transportation and real estate sectors, account for 81% of total carbon emission in China [31], three of which (except energy) have been heavily invested in megacities by business fixed investment in infrastructure and real estate which extensively use diesel vehicles with serious exhaust emissions. Thus, investment on fixed assets should be one of key sources linked with carbon emission and directly affects on air quality in the above cities. As tourism is recognized as a momentous source of income for megacities [5] and assesses the operation efficiency of tourism sector, this research takes tourism revenue as the proxy variable of tourism resource endowment [28] and the total number of star hotels [32]. The possible effects of foreign exchange rate between US dollar and RMB and CPI on spending of inbound tourists are considered as well.

In order to eliminate heteroscedasticity, those variables are measured by the natural logarithm. Table 1 provides the definition of variables.

Table 2 indicates the descriptive statistics for main variables. The average TAP in the sample is 4.785 with a standard deviation of 0. 35. The value gap between the maximum (5.799) and the minimum (3.931) is not wide. It implies that the air quality has not been significantly improved. Meanwhile, two of key air pollution sources, PM10 and PM2.5, have the mean value of 4.195 and 3.763 with the standard deviation of 0.34 and 0.41, respectively.

## 4. Model, Results and Robustness Test

### 4.1. Model

This study employs the following equation to test the correlation between NIT and TAP after controlling the five factors that may affect NIT in given cities by applying the fixed effects (FE) regression model.

The characteristic of the fixed effect model is reflected by adding n-1 dummy variables into the linear regression model, so that each section has a constant term. The changes of panel data are associated with individuals rather than with time in the regression of fixed effect [33]. Furthermore, this model just compares the selected groups of variables, and the conclusion is only applicable for the specific categories of each independent variable to be considered in the study [34]. This research aims to assess the influence of air pollution on the size of inbound tourism, and the different component of inbound tourists between real tourists<sup>2</sup> and nominal travellers and their distribution in the sample megacities are not the random samples. Therefore, we use the fixed effect model of urban panel data for sample self-selection and empirical analysis.

**Table 1.** Definition of variables.

Variables	Definitions
<i>Dependent variables</i>	
The numbers of inbound tourist ( <i>NIT</i> )	Logarithm of the numbers of inbound tourist in three major tourist cities in China
<i>Independent variables</i>	
Total air pollution ( <i>TAP</i> )	Logarithm of total air pollution including PM2.5, PM10 and Sulfur dioxide of the sample cities
Particulate matter 2.5( <i>PM2.5</i> )	Logarithm of particulate matter 2.5 of the sample cities
Inhalable particulate matter 10( <i>PM10</i> )	Logarithm of inhalable particulate matter 10 of the sample cities
Sulfur dioxide ( <i>SD</i> )	Logarithm of sulfur dioxide of the sample cities
<i>Control variables</i>	
GDP per capita ( <i>AGDP</i> )	Logarithm of gross domestic product per capita
Tourism revenue ( <i>TR</i> )	Logarithm of tourism revenue referring to the income sum of tourists' consumption in different tourist areas obtained by tourism sector
Investment on fixed assets ( <i>IFA</i> )	Logarithm of investment on fixed assets including investment on infrastructure, real estate sector and renovation
Consumer price index ( <i>CPI</i> )	Logarithm of CPI on the basis of the last year's index
Foreign exchange rate ( <i>FER</i> )	Logarithm of foreign exchange rate between RMB and US dollar

**Table 2.** Descriptive statistics of main variables.

Variable	Mean	Std. Dev.	Minimum	Maximum
Ln( <i>NIT</i> )(ten thousand people)	3.713	0. 531	1.964	4.657
Ln( <i>TAP</i> ) (Density )	4.785	0. 349	3.931	5.799
Ln( <i>PM2.5</i> ) (Density)	3.763	0. 413	2.772	5.023
Ln( <i>PM10</i> ) (Density)	4.195	0. 337	3.401	5.170
Ln( <i>SD</i> ) (Density)	2.200	0. 511	0.693	7.202
Ln( <i>AGDP</i> ) (RMB billions)	11.079	0. 757	9.302	12.427
Ln( <i>IFA</i> ) (RMB ten billions)	17.427	0. 956	15.198	19.048
Ln( <i>TR</i> ) (RMB billions)	13.020	1. 782	9.144	16.297
Ln( <i>CPI</i> )	4.621	0. 011	4.599	4.644
Ln( <i>FER</i> )	6.496	0. 042	6.415	6.563

$$\begin{aligned} \text{Ln}(NIT_{i,t}) = & \alpha_0 + \beta_1 \text{Ln}(TAP_{i,t}) + \beta_2 \text{Ln}(AGDP_{i,t}) + \beta_3 \text{Ln}(IFA_{i,t}) + \beta_4 \text{Ln}(TR_{i,t}) \\ & + \beta_5 \text{Ln}(CPI_{i,t}) + \beta_6 \text{Ln}(FER_{i,t}) + \varepsilon_{i,t} \end{aligned}$$

<sup>2</sup>Tourists just for going sightseeing are identified as real tourists in this paper. Nominal tourists involve foreign visitors for the purpose of international business, transnational conference, service staffs and employees (BCSEs) which are recognized as "tourists" by Chinese bureaus of statistics. Source: National Bureau of Statistics, China. <https://data.stats.gov.cn/easyquery.htm?cn=C01>.

Where subscript  $i$  and  $t$  represent city and time separately;  $NIT_{it}$  represents the number of inbound tourists for city  $i$ th at time  $t$ ;  $TAP_{it}$  symbolizes the total air pollution in city  $i$ th at time  $t$ , in which contains particulate matter 2.5 (PM2.5), inhalable particulate matter 10 (PM10), and sulfur dioxide (SD), respectively. All of variables, including control variables from  $\beta_2$  to  $\beta_6$  in the equation, are measured by the natural logarithm.

#### 4.2. Results

Table 3 presents the regression results of the correlation between dependent and independent variables. There is a significant negative correlation between the total air pollution (TAP) and the numbers of inbound tourist (NIT) (coefficient = - 0.246;  $p < 0.001$ ). It means that a 1 % increase in TAP may lead to reduce 0.246% of NIT. In models (2), (3) and (4), PM2.5, PM10 and SD, each of them has a statistically and significantly negative association with NIT (coefficient = - 0.240;  $p < 0.001$ ),(coefficient = -0.251;  $p < 0.001$ ) and (coefficient = - 0.110;  $p < 0.05$ ). These results show that poor air quality, regardless what kind of source for air pollution, generally has a negative relationship with NIT. Therefore, the number of inbound tourists drops when the level of air pollution increases.

With regard to control variables that are considered in this study, the regression results report that the effects of GDP per capita (AGDP) and tourism revenue (TR) on NIT are positive and highly significant. The positive relationship could be the result of inbound tourist consumption as a part of GDP per capita and tourism revenue. However, the significant negative correlation of the investment on fixed assets (IFA) with NIT suggests that carbon emission during the investment processes plays a negative role in the inbound scale of foreign tourists. Two more control variables, FER and CPI, don't present statistical significant results with NIT.

The aforementioned findings confirm a negative relationship between air pollution and the number of inbound tourists during the sample period of 2015-2019, which is statistically significant at the level of 1%. More importantly, two sources of air pollution, PM10 and PM2.5, show serious impact on inbound tourists. As Table 3 displays that a 1% increase in PM10 or PM2.5 leads to reduce the number of inbound tourists (NIT) by 0.251% and 0.240%, correspondingly. The negative influence of PM 10 on sample cities is even more than PM2.5. Furthermore, IFA presents negative relationship with NIT. When investment on fixed assets increases by 1%, the number of tourists reduces by 0.168%. As carbon emission, including PM2.5, PM10 and SD, constitutes the main source of air pollution in the processes of IFA. Thus, as expected, the number of inbound tourists may continue to decline if carbon emission of IFA is not substantially improved. Simply, the higher the level of air pollution is, the less the number of tourists will be. On the other hand, the effects of AGDP and TR on NIT are positive and highly significant. It means that inbound tourism has a significant contribution to push up GDP per capita and tourism revenue as well.

Also, the VIF (variance inflation factor) in all of variables is less than two which illustrates that multicollinearity is very low.

**Table 3.** Regression results for the correlation between air pollution and inbound tourism.

Variables	(1)	(2)	(3)	(4)
<i>Control variables</i>				
InAGDP	0.245*** (6.98)	0.247*** (6.99)	0.245*** (6.95)	0.235*** (6.70)
InTR	0.199*** (11.06)	0.197*** (10.42)	0.201*** (11.21)	0.212*** (13.44)
InIFA	-0.168*** (-6.44)	-0.168*** (-6.45)	-0.168*** (-6.45)	-0.165*** (-6.40)
InFER	-0.193 (-0.32)	-0.199 (-0.32)	-0.132 (-0.22)	-0.638 (-1.01)
InCPI	-0.032 (-1.32)	-0.031 (-1.30)	-0.032 (-1.32)	-0.033 (-1.41)
<i>Independent variables</i>				
InTAP	-0.246*** (-3.46)			
InPM2.5		-0.240*** (-3.93)		
InPM10			-0.251*** (-3.41)	
InSD				-0.110* (-2.23)
Constant	6.596** (2.79)	6.055** (2.63)	6.381** (2.71)	9.058* (2.10)
Observations	180	180	180	180
Maximal VIF	1.27	1.30	1.26	1.32
Year	yes	yes	yes	yes

Notes: \*p<0.05, \*\*p<0.01, \*\*\*p<0.001; t-values are in parentheses.

#### 4.3. Robustness checks

To further demonstrate the reliability of regression results, we use the generalized method of moments (GMM) estimator, that is, the system-GMM estimator, for robustness checks. Comparing with static panel data models (fixed effects or random effects) or ordinary least squares (OLS) regressions, the system-GMM estimator has greater efficiency and less bias [30] which is also commonly used to test the robustness of panel data [35,36].

Table 4 shows that the results of regression coefficient between the core explanatory variables, TAP, PM2.5, PM10 and SD, and NIT are significantly negative relationship. The effect of IFA on NIT is also negative and highly significant. Moreover, the coefficients of GDP per capita (AGDP) and tourism revenue (TR) with NIT indicate the positive relationship. These results are consistent with the findings in Table 3.

**Table 4.** Robustness check results of SYS-GMM.

Variables	(1)	(2)	(3)	(4)
<i>Control variables</i>				
InAGDP	0.245*** (7.42)	0.247*** (7.32)	0.245*** (7.44)	0.235*** (7.68)
InIFA	-0.168*** (-6.04)	-0.168*** (-6.07)	-0.168*** (-6.04)	-0.165*** (-6.00)
InTR	0.199*** (10.49)	0.197*** (10.05)	0.201*** (10.28)	0.212*** (9.49)
InFER	-0.193 (-0.29)	-0.199 (-0.29)	-0.132 (-0.19)	-0.638 (-0.92)
InCPI	-0.032 (-1.27)	-0.031 (-1.25)	-0.032 (-1.26)	-0.033 (-1.34)
<i>Independent variables</i>				
InTP	-0.391*** (-3.56)			
InPM2.5		-0.332*** (-3.47)		
InPM10			-0.390*** (-3.62)	
InSD				-0.166* (-2.06)
Constant	6.102 (1.58)	6.025 (1.56)	5.538 (1.44)	9.058* (2.14)
Observations	180	180	180	180
r2	0.697	0.697	0.697	0.705
Year	yes	yes	yes	yes

Notes: \*p<0.05, \*\*p<0.01, \*\*\*p<0.001; t-values are in parentheses.

In addition, we use sub-sample regression analysis for the sake of comparing the impact of two major air pollutants, PM2.5 and PM10, on inbound tourism in sample

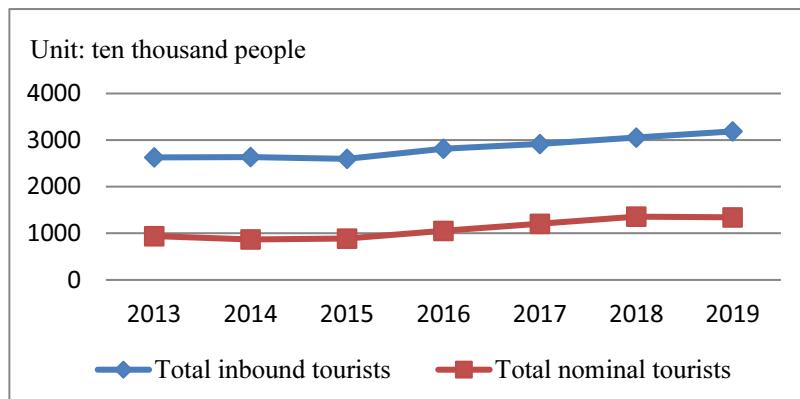
megacities. Table 5 shows the results of sub-sample regression about the correlation between air pollution and inbound tourism in Chongqing, Beijing and Guangzhou. The results indicate a significant negative correlation between inbound tourists and PM2.5 (coefficient = -0.327;  $p < 0.001$ ) and PM10 (coefficient = -0.403;  $p < 0.001$ ) in Chongqing, but the negative correlation is not significant in Beijing. Moreover, it seems that inbound tourists in Guangzhou are not affected by air pollution since the coefficient between inbound tourists and PM2.5 and PM10 become positive correlation without statistically significant.

**Table 5.** Regression results of sub-sample

	(1) Chongqing LnNIT	(2) Beijing LnNIT	(3) Guangzhou LnNIT
<i>Control variables</i>			
InAGDP	-0.180**(-2.86)	-0.199*(-2.63)	-0.126(-1.79)
InTR	0.277***(9.35)	-0.043(-0.70)	0.292*** (4.20)
InIFA	0.018(0.36)	0.195(1.33)	0.096(1.96)
InFER	-0.942(-1.09)	-0.627(-1.23)	0.735(1.20)
InCPI	-0.022(-0.73)	-0.008(-0.53)	0.020(0.64)
<i>Independent variables</i>			
InPM2.5	-0.327*** (-3.49)	-0.001 (-0.02)	0.001 (0.73)
InPM10		-0.403*** (-3.55)	-0.014 (-0.21) 0.002 (1.39)
Constant	7.088 (1.51)	8.667 (1.80)	7.971 (1.76) 8.161 (1.79) -2.962 (-0.94) -2.763 (-0.89)
Observations	60	60	60
F	64.60	65.09	5.267 5.256 8.224 8.668
r2	0.880	0.881	0.373 0.374 0.482 0.495

Notes: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ;  $t$ -values are in parentheses.

The reasons may result from the composition and distribution differences of inbound tourists in sample megacities. In fact, a large number of foreign visitors for the purpose of international business, transnational conference, service staffs and employees (BCSEs), that is, nominal tourists, account for about one third of the total number of inbound tourists in China (see Figure 1).



**Figure 1.** A comparison between the total number of inbound tourists and nominal tourists (2013-2019).  
Source: National Bureau of Statistics, China. <https://data.stats.gov.cn>.

Clearly, Beijing is a center of political, economic and cultural, while Guangzhou is the other heavyweight of economic and cultural where holds the fairs of export and import twice a year. Therefore, both of megacities attract a lot of foreign government

officials, employees and business men whom are not real tourists but have to intensively stay or travel in these cities. Those groups of BCSEs that are recognized as "tourists" by local bureaus of statistics would not be too care about the influence of regular air pollution when they come to Beijing or Guangzhou, even though Guangzhou ranks the second of carbon emissions in global urban [37].

One more puzzle of the results of sum-sample regression is that the regression coefficients on IFA are no statistically significant and have positive correlation with NIT. By contrary, the coefficients on AGDP are significant negative correlation with NIT in Beijing and Chongqing. The possible explanations are as follows:

First, when the control variables IFA are analyzed as independent variables with other variables, the explanatory power of IFA is diluted by other control variables [33].

Second, the main components of GDP (Y) include investment (I), consumption (C) and net exports (NX), that is,  $Y = I + C + NX$  [38]. Among the three component factors, some studies have proved that international trade and consumption don't reduce the size of inbound tourists despite of increasing air pollution [39-41]. IFA is a part of GDP and has dual effects on inbound tourism, expanding the size of inbound tourists by improving tourism infrastructure but decreasing the numbers of foreign tourists due to its carbon emission, for example. Thus, the negative correlation between AGDP and NIT, actually, shows the negative relationship between IFA and NIT.

Third, although the coefficient on AGDP is no significant with NIT in the case of Guangzhou, it should not be recognized that there is no significant negative relationship between IFA and NIT. As previously mentioned, a number of the BCSEs to be classified as tourists (nominal tourists) from different countries, come to Guangzhou for business fairs no matter what air quality is. This special phenomenon does not imply zeroing impact of air pollution on the health of nominal tourists and residents for sure.

The above reasons illustrate that the results of core independent variables and main control variable in Table 5 are no substantial difference from the regression results of benchmark in Table 3.

## 5. Conclusions

This paper empirically analyzes the effect of air pollution on the number of inbound tourists in major megacities, Beijing, Guangzhou and Chongqing, China. The sample to be selected composes a panel data from January 2015 to December 2019 that allows us to analyze the negative influence of air pollution on the number of inbound tourists before the COVID-19 pandemic.

The empirical results demonstrate that the main air pollutants are linked with PM10 and PM2.5 owing to the combustion of coal, gasoline, fossil fuels and vehicle emissions in sample megacities. Especially, the most serious air pollution is connected with PM10. Furthermore, the effect of air pollution related to carbon emissions from business fixed investment on the number of inbound tourists is a negative correlation, while the influence of GDP per capita and tourism revenue on NIT reveal a positive relationship in accordance with the regression results of benchmark.

The different results between benchmark and sub-sample regressions present that air pollution indeed shrink the size of foreign real tourists but not on nominal ones whom concentrate in Beijing or Guangzhou for business and non-tour purposes. The results of sub-sample confirm the negative relationship between AGDP and NIT, and

thus provide an evidence about the same correlation between IFA and NIT in Beijing and Chongqing because the control variable IFA is a part of GDP. The exception of Guangzhou case stems from the special situation of nominal tourists. Consequently, there is no reason to deny the empirical results of benchmark by means of some coefficients without significant in sub-sample regression.

In order to stimulate the development of inbound tourism in Chinese megacities, it is essential to reduce carbon emissions associated with business fixed investment in which includes sectors of infrastructure, real-estate and motor vehicles, on fossil fuels based, so that effectively improve air quality from one of important original sources.

The fundamental strategies to lessen air pollution and flourish inbound tourism in megacities are to promote clean energy for green IFA through shifting infrastructure structure from fossil fuel consumption to new infrastructure represented by cloud computing, block-chain and artificial intelligence on information network based, spreading new building materials with green, ecological and environmental protection for real-estate sector, and updating engine, filter and fuel with low carbon emissions for clean diesel vehicles and other vehicles. During the processes of transforming energy type from high carbon fossil to renewable and low carbon, it is essential to popularize clean energy technologies with effective policy frame for radically improving air quality of megacities in China.

## **Author Contributions**

Nianlin Zhou wrote the manuscript; Yeli Gu analyzed the data; Manyuan Jiang modified the manuscript. All authors read and approved the final manuscript.

## **Funding**

This research was funded by the Education Bureau of Guangdong Province of China (Grant No. 2018GXJK272) and the research center of polar issues of South China Business College, Guangdong University of Foreign Studies (Grant No. NGJD 201901) and (Grant No. 20 JD004A).

## **Acknowledgments**

The authors acknowledge financial support from the Education Bureau of Guangdong Province of China, the research center of polar issues of South China Business College, Guangdong University of Foreign Studies, and editors and reviewers' valuable instructions for revision of this manuscript. The authors are responsible for all errors.

## **Conflicts of Interest**

The authors declare no conflict of interest.

## Note

- ① The term of megacities means that there are more than 10 million permanent residents living in the same urban areas. Source: The State Council of China, Notice on the adjustment of classification standard of urban scale, The State Council of China [2014] No.51, Oct. 29, 2014.

## References

- [1] Danish government. Pioneers in clean energy. 2021. Danish government website. Available online: <https://denmark.dk/innovation-and-design/clean-energy> (accessed on 5 March 2021).
- [2] Ministry of the Environment. Sweden's climate policy framework. Report of Swedish government. 2021 Mar; 11.
- [3] Xi JP. Work together to build a climate governance mechanism of win-win, fair and reasonable. Paris: Speech at the opening ceremony of the Paris conference on climate change. Xinhua News Agency. 2015 Dec; 1.
- [4] Wan J, Yan JP, Wang XM, Liu ZQ, Wang H, Wang T. Spatial-temporal pattern and its influencing factors on urban tourism competitiveness in city agglomerations across Guanzhong plain. Sustainability. 2019 Nov; (11):1-24.
- [5] Łapko A, Panasiuk A, Strulak-Wójcikiewicz R, Landowski M. The state of air pollution as a factor determining the assessment of a city's tourist attractiveness—based on the opinions of polish respondents. Sustainability. 2020 Dec; (12): 1-21.
- [6] Yang ZS, Su JH, Yang H. Exploring urban functional areas based on multi-source data: a case study of Beijing. Geography Research. 2021 Feb; (2): 477-94.
- [7] Deng TT, Liu S, Hu YK. Can tourism help to revive shrinking cities? An examination of Chinese case. Tourism Economics. 2021 May; (5):34-45.
- [8] Milewski D. Regionalne uwarunkowania rozwoju turystyki na. Przykładzie. 2019 Jul; (7): 55-68.
- [9] Zhou LQ, Bi J. The influence of cultural distance on international tourism destination choices: a case study of Chinese inbound tourism market. Journal of Zhejiang University. 2017 Apr; (4):130-42.
- [10] Sun GI, Qian Q. The impact of air pollution on the development of urban tourism in China . Journal of SCNU. 2020 Sep; (9): 65-73.
- [11] Ding L, Wu XG, Ding J. Construction and application of urban tourism competitiveness evaluation index system application. Economic Geography. 2006 Mar; (3): 511-525.
- [12] Łapko A, Panasiuk A. Water tourism as a recipient of transport Services on the example szczecin. Transportation Research. 2019 (39): 290–299.
- [13] Deng F, Yuan FY, Xu L, Li Z. Tourism, transportation and low-carbon city system coupling coordination degree: A case study in Chongqing municipality, China. International Journal of Environment Research and Public Healthy. 2020 Sep; (17):1-17.
- [14] Bieger T, Wittmer A. Air transport and tourism—Perspectives and challenges for destinations, airlines and governments. Journal of Air Transportation Management. 2006 Dec; (12): 40–56.
- [15] Becken S, Simmons DG, Frampton C. Energy use associated with different travel choices. Tourism Management. 2003 Dec; (24): 267–277.
- [16] Mohamad NS, Deni SM, Ul-Saufie AZ. Application of the first order of Markov chain model in describing the PM10 occurrences in Shah Alam and Jerantut, Malaysia, Pertanika. Journal of Science Technology. 2018 Dec; (26):138-152.
- [17] Lin CA, Chen YC, Liu CY, Chen WT, Seinfeld JH, Chou CCK. Satellite-derived correlation of SO22019, NO2, and aerosol optical depth with meteorological conditions over East Asia from 2005 to 2015. Remote Sensing. 2016 Nov; (11): 1738.
- [18] Signals EEA. Towards clean and smart mobility. Transport and environment in Europe. European Environment Agency. 2015 Jun; (6): 12-26.
- [19] Li L, Liu Y. Study on public negative externality of air quality on inbound tourism in megacities. Ecology Economics. 2019 Oct; (10): 124-139.
- [20] Zhang C, Gao AJ, Ding PY. Impacts of haze on the intention to visit china of overseas tourists: based on destination image and risk perception theories. Tourism Tribune. 2017 Dec; (12): 58-67.
- [21] Beckens S, Jin X, Zhang C et al. Urban air pollution in China: destination image and risk perceptions. Journal of Sustainable Tourism. 2016 May; (5):1-14.

- [22] Saenz DMO, Rosselló J. Modeling tourism impacts on air pollution: The case study of PM10 in Mallorca. *Tourism Management*. 2014 Oct; 40(10): 273-281.
- [23] Chen SY, Lin BQ. Research status and prospect of economics of energy, environment and climate change in China - a summary of the first forum of Chinese economists on energy, environment and climate change. *Economic Research*. 2019 Jul; (7): 203-218.
- [24] Jiang L, He SX, Cui YZ. Analysis of sulfur dioxide pollution control in China: an empirical study based on satellite observation data and spatial metrology model. *Journal of Environment Science*. 2021 May; (3):1153-1164.
- [25] Sun CW, Luo Y, Yao X. The effects of transportation infrastructure on air quality: evidence from empirical analysis in China. *Economic Research*. 2019 Aug; (8):136-151.
- [26] Shao S, Li XB, Cao JH. Urbanization promotion and haze pollution governance in China. *Economic Research*. 2021 Feb; (2):148-164.
- [27] Tanpaifang. An effective way to reduce pm2.5.source: controlling carbon emissions in China. 2014. Carbon emissions trading network of China. Available online: <http://www.tanpaifang.com> (accessed on 11 April 2021)
- [28] Liu JY, Chen YP, Xia X. Research on the effect of air pollution on the development of inbound tourism in China. *Research Science*. 2018 Jul; (7): 1473-1482.
- [29] Blanca F, Caballero J, Gonzalez RM, Lozano OM, Perez F. Goal programming synthetic indicators:an application for tourism in Andalusian coastal counties. *Ecological Economics*. 2010 Feb; (2): 158-172.
- [30] Academy of Tourism of China. Annual report on inbound tourism development in China. Beijing: The report of Academy of Tourism of China. 2014. 25-79 pp.
- [31] Duan GS. Carbon neutral driving millions investment. *Security Daily*. 2021 May 22.
- [32] Sun GJ, Qian Q. The impact of air pollution on the development of urban tourism in China . *Journal of SCNU*. 2020 Sep; (9): 65-73.
- [33] Lian YJ. Research on investment efficiency of Listed Companies in China. *Press of Economic Management*. 2009 Aug; (8): 96-102.
- [34] Pan SC. *Econometrics*. Beijing: Press of Renmin University of China; 2009. 112-132 pp.
- [35] Faulkender M, Flannery MJ, Hankins KW, Smith JM. Cash flows and leverage adjustments. *Journal of Financial Economics*. 2012 Jan; 103(1): 632-646.
- [36] Warr RS, Elliott WB, Koeter KJ, Oztekin O. Equity mispricing and leverage adjustment costs. *Journal of Financial Quantitative Analysis*.2012 Sep; 47 (9): 589-616.
- [37] Azam M, Khan AQ, Bin AH. The impact of CO<sub>2</sub> emissions on economic growth: evidence from selected higher CO<sub>2</sub> emissions economies. *Environmental Science and Pollution Research*. 2016 Jul; (7): 6376-6389.
- [38] Mankiw NG. *Essentials of Economics*. Boston: Press of Harvard University; 2002. 278 p.
- [39] Wang JJ. The Empirical analysis on the relationship between inbound tourism and import & export trade. *Journal of Economic Issues*. 2012 Nov; (11): 99-103.
- [40] Mi X, Liu ZY. Interactions and regional differences between inbound tourism and international trade in Xinjiang. *Journal of L Y Normal University*. 2021 Apr; (4):12-18.
- [41] Su JJ, Sun GN, Xu ZY. The pull effect of tourism development on demands for investment, consumption and exports in China. *Tourism Tribune*. 2014 Feb; (2): 25-35.

# Research on and Application of Tunnel Structure Defects Prediction Using Machine Learning Methods

Bo SHI<sup>a,b,c,d</sup>, Hui SU<sup>a,c,d</sup>, Xu DU<sup>a,c,d</sup>, Bao JIAO<sup>a,c,d</sup>, Lin WANG<sup>a,c,d,e<sup>1</sup></sup>

<sup>a</sup>SGIDI Engineering Consulting (Group) Co., Ltd, Shanghai 200093, China

<sup>b</sup>School of Computer Science and Technology, Fudan University, Shanghai 2000438, China

<sup>c</sup>Shanghai Engineering Research Center of Geotechnical Test for Underground Space, Shanghai 200093, China

<sup>d</sup>Shanghai Professional Technology Service Platform of Geotechnical Engineering, Shanghai 200093, China

<sup>e</sup>Department of Geotechnical Engineering, College of Civil Engineering, Tongji University, Shanghai 200092, China

**Abstract.** With the rapid development of underground engineering in China, more metro tunnels are being constructed, the mileage of subway tunnels is increasing, and the corresponding problems of tunnel structure diseases are becoming more prominent. At present, the treatment of tunnel structural diseases mainly relies on manual inspection and identification, and research on defects prediction is still lacking. Because of the complexity of the factors affecting tunnel structure diseases, it is difficult to analyze the causes and development trend of the diseases comprehensively by manual analysis. Fortunately, machine learning methods have gained popularity in classification and regression tasks in recent decades. Many algorithms, such as decision tree algorithms, the random forest algorithm, and XGBoost, have been applied in fields including finance, engineering, and transportation. This study aimed to analyze the prediction effect of machine learning models by feeding 68055 segment lining rings of six subway lines in a city. According to the disease records from 2014 to 2016 and the corresponding convergence and characteristic data, defect conditions in 2017 were predicted and compared with real defect conditions in 2017. The accuracy rates and F1 values of the predicted results were all above 80%. The prediction results can help tunnel maintenance departments and relevant government regulators make auxiliary decisions to control tunnel structure diseases, and can help them focus on the tunnel interval of severe diseases to clarify the development trend of tunnel disease.

**Keywords:** Tunnel engineering, operational tunnel structure safety, tunnel structure disease prediction, machine learning

## 1. Introduction

With the rapid development of China's economy and the continuous acceleration of urban construction, underground engineering is in a stage of rapid development, and the

---

<sup>1</sup>Corresponding author: Lin Wang, SGIDI Engineering Consulting (Group) Co., Ltd, Shanghai 200093, China; E-mail: 1410024@tongji.edu.cn.

corresponding tunnel construction is also increasing. Increased operation time, coupled with the influence of tunnel construction conditions, the surrounding tunnel environment, and other factors, have made various tunnel diseases more prominent, especially in soft soil areas, such as Shanghai, Hangzhou, and other cities in China. Owing to the particularity of soft soil, leakage, segment damage, cracks, and other situations are more significant. Serious tunnel structure damage can lead to severe accidents, thereby affecting the safe operation of the subway and seriously threatening the safety of life and property.

However, tunnel structure safety is affected by many factors, such as geology, design, construction, and operation, and the causes of diseases are complex [1]. Yuan et al. defined a quantitative assessment of the service states of the structure in an operational tunnel and proposed a framework of assessment procedure. Previous studies on tunnel assessments mostly used data mining technology to evaluate the state of tunnel structure and put forward corresponding measures [2]. Alimoradi et al. proposed that an appropriately trained neural network could reliably predict weak geological zones in front of a tunnel face accurately [3]. Adoko et al. used Multivariate Adaptive Regression Spline (MARS) and ANN models to predict the diameter convergence of a high-speed railway tunnel in weak rock and concluded that MARS constitutes a reliable alternative to ANN in modeling geo-engineering problems [4]. Mahdevari et al. developed a dynamic model based on Support Vector Machines (SVMs) for the prediction of convergence in the tunnel, and got good results in situ measured ones [5]. Feng et al. proposed a Bayesian approach to improve time-dependent convergence predictions and the results showed that the proposed method could help improve the accuracy of predictions and reduce their uncertainty [6].

However, previous studies concerned with the deformation prediction of the ring segments such as convergence, and these studies lacked a forecast of the development trend of tunnel structure defects based on a large amount of engineering data. Unlike predicting the deformation in tunnels in previous studies, the paper first provided an idea for predicting defects and defect types using machine learning methods. At present, the tunnel maintenance department in China has accumulated a large amount of disease information according to regular tunnel inspection. Unfortunately, the use of the data is limited to basic inquiry and statistical work, so more intelligent methods are needed to predict the development trend of the disease. Accurately and efficiently predicting possible tunnel diseases has important engineering guiding significance to help tunnel maintenance departments determine the inspection scope and avoid major disasters.

In this study, three machine learning methods, the decision tree, random forest, and XGBoost, were applied to the prediction of tunnel structure diseases to help the tunnel maintenance department understand the development trend of the disease and make timely decisions on prevention and control measures.

## **2. Methodology**

### *2.1. Decision Tree*

Decision tree is a machine learning prediction model that represents a mapping relationship between features and target values. It is a decision analysis method that is used to obtain the expected probability so as to evaluate risk and judge its feasibility by forming a decision tree based on the known probability of various situations. There are

several decision tree algorithms, including Iterative Dichotomiser 3 (ID3), C4.5 and Classification and Regression Tree (CART) [7-9]. Its basic principle is as follows.

For data feature vectors  $x_i \in R^n, i=1, \dots, l$  and a label vector  $y \in R^l$ , a decision tree recursively partitions the feature space, such that the samples with the same labels or similar target values are grouped together. A node,  $m$ , can be represented by  $Q_m$  with  $N_m$  samples and  $Q_m$  is the subset of the whole dataset. For each candidate split,  $\theta = (j, t_m)$ , consisting of a feature,  $j$ , and threshold,  $t_m$ , partition the data into  $Q_m^{left}(\theta)$  and  $Q_m^{right}(\theta)$  subsets, which are calculated as follows:

$$Q_m^{left}(\theta) = \{(x, y) | x_j \leq t_m\} \quad (1)$$

$$Q_m^{right}(\theta) = \{(x, y) | x_j > t_m\} \quad (2)$$

If a target is a classification outcome taking on values  $0, 1, \dots, K-1$  for node  $m$ , we can calculate  $p_{mk}$  (equation (3)) as the proportion of class  $k$  observations in node  $m$ . Specially, function  $I(y=k)$  in equation (3) is called the Indicator function, which can be calculated as in equation (4).

$$p_{mk} = 1/N_m \sum_{y \in Q_m} I(y=k) \quad (3)$$

$$I(y=k) = \begin{cases} 1 & \text{if } y=k \\ 0 & \text{if } y \neq k \end{cases} \quad (4)$$

Subsequently, common measures of impurity (loss function), including ‘Gini index’ in equation (5) and ‘Entropy’ in equation (6) [10], are calculated as follows:

$$H(Q_m) = \sum_k p_{mk} (1 - p_{mk}) \quad (5)$$

$$H(Q_m) = -\sum_k p_{mk} \log(p_{mk}) \quad (6)$$

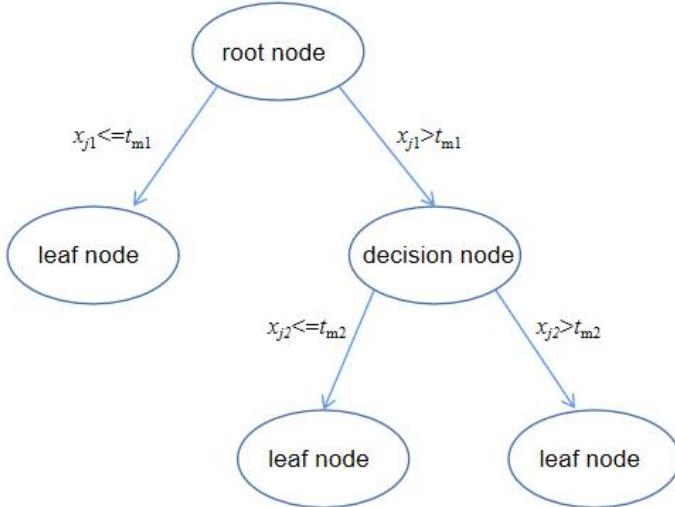
Then, the quality of a candidate split of node  $m$  can be computed using the loss function (equation (7)) [10].

$$G(Q_m, \theta) = \frac{N_m^{left}}{N_m} H(Q_m^{left}(\theta)) + \frac{N_m^{right}}{N_m} H(Q_m^{right}(\theta)) \quad (7)$$

Finally, choose proper parameters to minimize the loss function in equation (8), and recurse for subsets  $Q_m^{left}(\theta^*)$  and  $Q_m^{right}(\theta^*)$  until the node cannot be split or the given maximum depth is reached.

$$\theta^* = \arg \min_{\theta} G(Q_m, \theta) \quad (8)$$

The sketch map of a decision tree is shown in Figure 1. For a given dataset, it can be recursively split into ‘tree’ as the sketch map.



**Figure 1.** Sketch map of a decision tree.

## 2.2. Random Forest

Random forest [11,12] is an ensemble method that combines several base estimators to improve accuracy and stability over a single estimator. Furthermore, the random forest algorithm builds several estimators independently and averages their predictions, which is called the bagging method and can reduce the variance of a base estimator by introducing randomization into its construction procedure [13,14,15]. In this study, CART was used as the base estimator, and “Gini” was used as the loss function.

The process of the random forest algorithm is as follows. Firstly, extract  $m$  training samples from the original data set using the Bootstrap method;  $n$  sub training sets are obtained by carrying out put-back sampling  $n$  times. Secondly, select  $p$  attributes from  $m$  attributes on each training set to generate  $n$  CART decision trees. Thirdly,  $n$  CART decision trees are formed into a random forest. For the classification problem, the majority voting method is adopted, and the final classification result is determined by voting according to the classifier of  $n$  trees.

## 2.3. XGBoost

XGBoost is also an ensemble method that combines several base estimators [16]. However, unlike the random forest algorithm and bagging methods, XGBoost is a boosting method, and the base estimators are built sequentially to reduce the bias of the combined estimators.

The regularized objective function is as follows:

$$L(\phi) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k), \text{ s.t. } \Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (9)$$

Here,  $l$  is a differentiable convex loss function that measures the difference between the prediction,  $y_i$ , and the target,  $\hat{y}_i$ . The second term,  $\Omega$ , penalizes the complexity of the model, which can help avoid over-fitting. Each  $f_k$  corresponds to an independent tree structure.

When we predict the  $i$ -th instance at the  $t$ -th iteration, the prediction can be written as:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x) \quad (10)$$

Then, we minimize the following objective:

$$L^{(t)} = \sum_{i=1} l\left(y_i, (\hat{y}_i^{(t-1)}) + f_t(x_i)\right) + \Omega(f_t) \quad (11)$$

Second-order approximation can be used to quickly optimize the objective above. Then, we can get the following objective:

$$L^{(t)} = \sum_{i=1} \left[ l\left(y_i, (\hat{y}_i^{(t-1)})\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \quad (12)$$

where  $g_i$  and  $h_i$  are first- and second-order gradient statistics on the loss function.

Obviously, the regularization term is introduced to control the complexity of the model and avoid over-fitting, and we can define the objective function as needed.

### 3. Model Construction and Application

In this study, data of 68,055 segment lining rings of six subway lines in a city were fed into the machine learning models mentioned above. According to defects records from 2014 to 2016 and corresponding convergence and characteristic data, defects conditions in 2017 were predicted and compared with real defects conditions in 2017 to analyze the prediction effect of the models.

#### 3.1. Data Preprocessing

In machine learning field, the data determines the upper limit of machine learning, and the algorithm is just as close to that limit as possible. Thus, the selection of features is very important for machine learning models. Fortunately, although tunnel structure defects are affected by multiple factors, we can get the features base on engineering experience.

Combined with engineering experience, various factors including convergence, buried depth, initial convergence, PS value and sensitivity value of the soil layer, geological division, hydrogeological division, cover condition and open time were determined and corresponding data were collected (as shown in Table 1, Table 2, and Table 3).

**Table 1.** Sample data of tunnel segments' time-invariant features.

ring_id	263517	260196	260012	148570	321966	9535	207927
ps_value	0.4679	0.7598	0.599	0.6995	0.5012	0.5677	1.6795
buried_depth(m)	8.103	8.655	6.714	14.201	9.071	11.842	24.258
init_convergence(m)	0.0200	0.0183	0.0276	0.0256	0.0347	0.034	0.0280
Sensitivity_value	4.4738	1.9422	2.0805	4.3165	2.5142	2.5011	1.9164
open_time(year)	1993	1993	1993	2010	2007	2010	2013
Relationships_with_bypass	A	E	A	E	E	E	E
cover_condition	road	road	park	water	yard	road	buiding
geological_division	IA	IIB	IIB	IIA	IIIAE	IIIBE	IIIB
hydrogeological_division	first	fourth	fourth	third	seventh	fifth	sixth

**Table 2.** Sample data of tunnel segments with diseases.

Ring number	Defect types	Discovery date of defect
263517	leakage	2014
263517	leakage	2017
263518	leakage	2014
263518	leakage	2017
263592	leakage	2014
263736	cracking	2017
264061	leakage	2014
264064	leakage	2015
264139	spalling	2014
264139	spalling	2016
264139	leakage	2016

**Table 3.** Sample data of tunnel segments' time-variant features (inner diameter and deformation against inner diameter).

Ring number	Measure time	Inner diameter(m)	Deformation(m)
263517	2014	5.55417	0.05417
263517	2015	5.55218	0.05218
263517	2016	5.5514	0.0514
263517	2017	5.5595	0.0595
263518	2014	5.55417	0.05417
263518	2015	5.55218	0.05218
263518	2016	5.5514	0.0514
263518	2017	5.5595	0.0595
263521	2014	5.54813	0.04813
263521	2015	5.54602	0.04602
263521	2016	5.547	0.047
263521	2017	5.5467	0.0467

Before training machine learning models, the data need to be preprocessed, and continuous variables need to be normalized. Normalization refers to limiting features' data within the range of 0 to 1. The calculation is as follows:

$$X = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (13)$$

where  $X_i$  denotes structural feature data before normalization,  $X_{\max}$ ,  $X_{\min}$  denote the maximum and minimum values of the feature, respectively; and  $X$  denotes the normalized structural feature data.

Categorical variables are processed with One-Hot Encoding. One-Hot Encoding uses 0 and 1 to represent parameters and N status for a state code. For example, tunnel segment lining rings in the geological division zone can be divided into four district, and the first district can be expressed as [1,0,0,0] and the second district can be expressed as [0,1,0,0], the third district and fourth district can be expressed as the same way.

In the paper, convergence, buried depth, initial convergence, open time, PS value and sensitivity value of the soil layer were calculated as continuous variables and geological division, hydrogeological division, cover condition were calculated as categorical variables.

### 3.2. Model Training and Evaluation

After data processing, the data were divided into a training set (data from 2014–2016) and test set (data from 2017). Then the training set was fed into Decision Tree model, Random Forest model and XGBoost model for training and the test set was used for evaluating the models. Additionally, we used the module called GridSearchCV [17] from scikit-learn to tune the hyper-parameters of the three machine learning models.

Model evaluation generally includes accuracy rate, recall rate, and F1 value in classification tasks. The confusion matrix has the following four markers in binary classification in Table 4. True Positive (TP) means predicting a positive sample as a positive sample; False Negative (FN) means predicting a positive sample as a negative sample; False Positive (FP) means predicting a negative sample as a positive sample; and True Negative (TN) means predicting a negative sample as a negative sample.

**Table 4.** Four markers in binary classification results.

Predict (column) / label (row)	Positive	Negative
Positive	TP	FP
Negative	FN	TN

The precision rate is the ratio of the correct prediction among predicted positive samples. It can be calculated as follows:

$$Precision = \frac{TP}{TP + FP} \quad (14)$$

The recall rate is the ratio of the correct prediction among positive label samples. It can be calculated as:

$$Recall = \frac{TP}{TP + FN} \quad (15)$$

The F1 value is the comprehensive evaluation of precision rate and recall rate. It is the harmonic average value of the two values, and its calculation is as follows:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (16)$$

Receiver operating characteristic curve (ROC curve) is another way to evaluate machine learning model. For a ROC curve of machine learning model, x axis denotes the False Positive rate and y axis denotes the True Positive rate. As the False Positive rate is only related to negative samples and the True Positive rate is only related to positive samples, ROC curve doesn't change as the class distribution (the ratio of the positive samples and the negative samples) changes, that is, it is very suitable for evaluating models of imbalance dataset. AUC is the area under ROC curve and it shows the quality of a classifier.

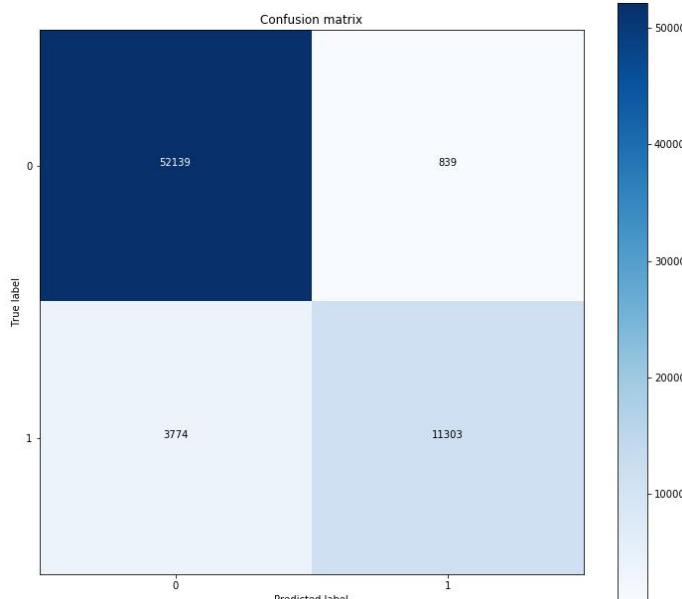
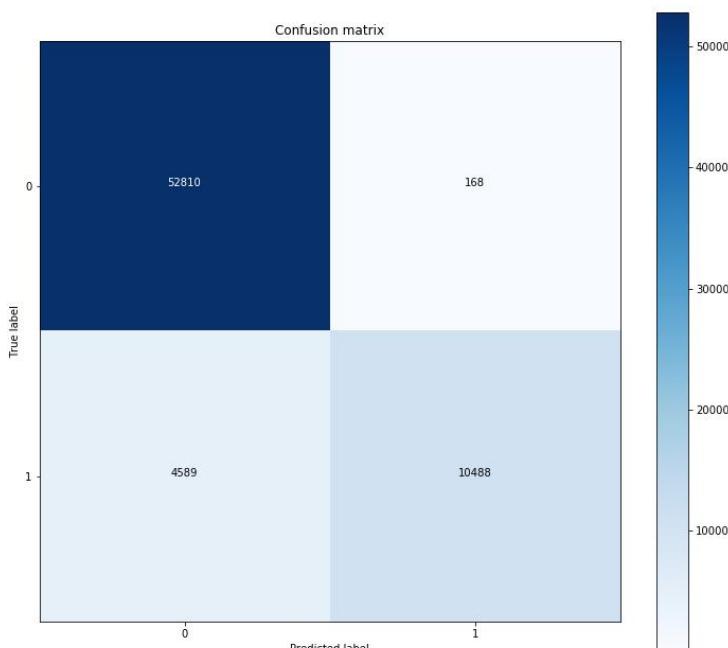
### 3.3. Results and Discussion

As most of the tunnel segment lining rings were free of disease, the accuracy could not truly reflect the effect of the model. Therefore, we randomly selected a subset of instances with no defect to match the number of instances with defects to reduce the imbalance of the dataset and used four indexes, accuracy, precision, recall, and F1 values, to comprehensively evaluate the effect of the model. We compared the real condition of tunnel segment defects with the predicted condition to generate the confusion matrix (Figures 2–4), which represents the effect of the models. We also plotted the ROC curves of the three models and the corresponding AUC values were calculated in the Figure 5.

In this study, the segment lining rings' diseases were set as positive samples. In Figures 2–4, the horizontal axis denotes the label of predicted value, and the ordinate denotes the label of true value, in which 0 means no disease and 1 means disease. As the evaluation values in Table 5 show, the training model worked well. The results of the three models were very close, and all demonstrated quite accurate predictions on the tunnel diseases; the accuracy rates were all above 90%, and the F1 values were all above 80%. As we can see, we have got high accuracy rate and the models still leave room for improvement according to the results of recall rate. From ROC curves and AUC values in Figure 5, we can see that all models got high AUC values which means the algorithms we used are robust though the dataset is imbalanced. Furthermore, we can see that XGBoost model and Random Forest model are better than Decision Tree model as the ensemble learning methods are superior to a single estimator.

**Table 5.** Evaluation indexes of the machine learning models.

Algorithms	Accuracy	Precision	Recall	F1 value
Decision tree	93%	98.4%	74%	83%
Random forest	93%	98.4%	69.5%	81%
XGBoost	93%	99.8%	69%	81.4%

**Figure 2.** Prediction results of Decision Tree model of binary classification (confusion matrix).**Figure 3.** Prediction results of Random Forest model of binary classification (confusion matrix).

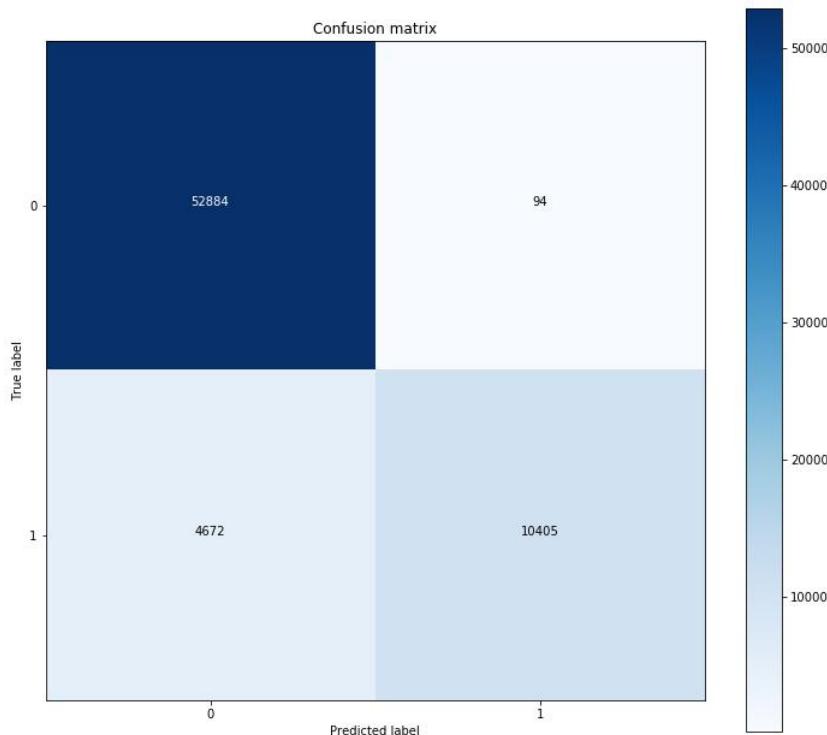


Figure 4. Prediction results of the XGBoost model of binary classification (confusion matrix).

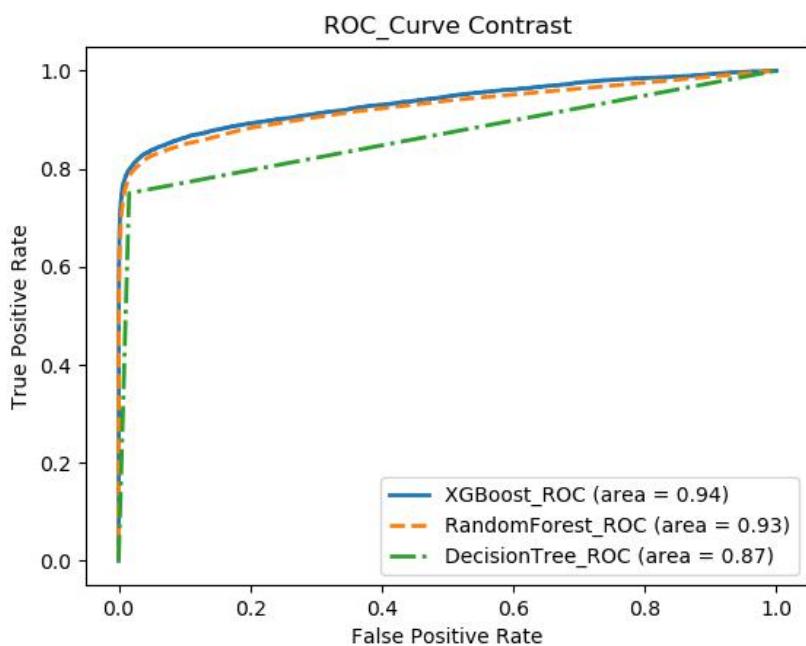


Figure 5. ROC curve and AUC values of the three models.

#### 4. Conclusion

Given the complexity of the influencing factors of metro tunnel structural diseases, the study used three machine learning methods to predict the occurrence of the defects in metro tunnels. Through the three models, we predicted tunnel structural defects and obtained satisfactory results. The accuracy rates were all above 90%, and the F1 values were all above 80%. It is important to note that it is nearly impossible to achieve completely accurate prediction using machine learning methods. Furthermore, ROC curve of the three models were plotted and AUC values were calculated. The AUC values of XGBoost model, Random Forest model, Decision Tree model are 0.94, 0.93 and 0.87 respectively which means that the algorithms we chose are robust for imbalanced engineering data. Overall, our proposed method can be useful to aid tunnel maintenance departments owing to the high prediction accuracy rate.

The prediction results can provide auxiliary decision-making assistance for tunnel maintenance departments and relevant government regulatory departments to prevent and control tunnel structural diseases and focus on tunnel sections where serious diseases may occur so as to further clarify the development trend of tunnel diseases. This study provides a new idea for the prediction of tunnel disease, but the current research can only give the probability of the occurrence of the disease (binary classification). Subsequent research can be extended to the prediction of the occurrence of the disease (multiple classification) and its probability, which will provide more accurate auxiliary guidance for tunnel maintenance and supervision authorities.

#### Acknowledgements

The authors would thank Social Development Projects of Shanghai ‘Science and Technology Innovation Action Plan’(18DZ1205900,19DZ1200801) and National Natural Science Foundation of China (42002272) for support.

#### References

- [1] Chen X, Zhu H, Li X, Lin X and Wang X. Probabilistic performance prediction of shield tunnels in operation through data mining. *Sustainable Cities and Society*. 2019; 44: 819-29.
- [2] Yuan Y, Bai Y and Liu J. Assessment service state of tunnel structure. *Tunn. and Undergr. Space Technol.* 2012; 27: 72-85.
- [3] Alimoradi A, Moradzadeh A, Naderi R, Salehi M Z and Etemadi A. Prediction of geological hazardous zones in front of a tunnel face using TSP-203 and artificial neural networks. *Tunn. and Undergr. Space Technol.* 2008; 23:711-7.
- [4] Adoko A, Jiao Y, Wu L, Wang H and Wang Z. Predicting tunnel convergence using Multivariate Adaptive Regression Spline and Artificial Neural Network. *Tunn. and Undergr. Space Technol.* 2013; 38:368-76.
- [5] Mahdevari S, Hamid S and Seyed R. A dynamically approach based on SVM algorithm for prediction of tunnel convergence during excavation. *Tunn. and Undergr. Space Technol.* 2013; 38: 59-68.
- [6] Feng X, Jimenez R, Zeng P and Senent S. Prediction of time-dependent tunnel convergences using a Bayesian updating approach. *Tunn. and Undergr. Space Technol.* 2019; 94: 103-118.
- [7] Breiman L, Friedman J, Olshen R, and Stone C. Classification and regression trees. Wadsworth, Belmont, CA. 1984.
- [8] Quinlan J. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers. 1993.
- [9] Hastie, Tibshirani R and Friedman J. Elements of Statistical Learning. Springer. 2009.
- [10] Quinlan J. Introduction of decision trees. *Machine Learning*. 1986;1:81-106.
- [11] Breiman L. Bagging predictors. *Machine Learning*. 1996;24 :123-40.

- [12] Breiman L. Random forests Machine Learning. 2001;45 5-32.
- [13] T Ho. The random subspace method for constructing decision forests. Pattern Analysis and Machine Intelligence. 1998;20: 832-44.
- [14] Pedregosa et al. [Scikit-learn: Machine Learning in Python](#). JMLR. 2011; 12:2825-30.
- [15] G Louppe and P Geurts. Ensembles on random patches. Machine Learning and Knowledge Discovery in Databases. 2012; 346-61.
- [16] Chen T, Carlos G. XGBoost: A scalable tree boosting system [C]. SIGKDD. 2016.
- [17] Buitinck et al [API design for machine learning software: experiences from the scikit-learn project](#) ECML PKDD Workshop: Languages for Data Mining and Machine Learning. 2013;108-22.

# Evaluation of Water Resources Sustainable Utilization in Wuhan Based on the Model of SUDE

Lan DONG<sup>1</sup>

*Evergrande School of Management, Wuhan University of Science and Technology,  
Wuhan 430080, China*

**Abstract.** SUDE model is a systematic sustainable use model based on the view of the whole life cycle of water resources. The analysis of the model water resources has been divided into four stages according to the state of motion water resources at different times in its life cycle, namely the water supply stage ("S"), water use phase ("U"), water collection, treatment and drainage stage ("D") and water return to natural ecological (ecology) environment for dilution, degradation stage ("E"). On the analysis above, SUDE theoretical model of water system has been built and corresponding evaluation index system has therefore been established. Then, we use entropy weight TOPSIS method to evaluate the sustainability of water resource during 2013~2017 years in Wuhan, and finally the corresponding conclusions and suggestion have been drawn according to the result of evaluation in the paper.

**Keywords:** SUDE model, water resources, sustainability, entropy weight TOPSIS method

## 1. Introduction

Water system is not only the resource guarantee of a region's economic development, but also an important part of the natural ecological environment. The degree of its healthy and effective utilization is an important embodiment of the harmony between human society and nature, which directly affects the whole social and economic sustainability [1]. Therefore, it is an important element of sustainable development in a region.

The meaning of sustainable development is to meet the needs of the present generation and not to endanger the ability of future generations to meet their needs. The sustainable use of water resources is defined as the whole process of supporting the coordinated development of population, resources, environment and economy and meeting the water needs of both within and between generations, while maintaining water sustainability and ecosystem integrity [2].

## 2. Relevant Literature Review

---

<sup>1</sup>Corresponding author: Lan Dong, Evergrande School of Management, Wuhan University of Science and Technology, Wuhan 430080, China; E-mail: strawberry\_d@163.com.

The research of sustainable water resources utilization in management concept and quantitative evaluation have been carried out. Koop (2015), Kasim (2014) believed that the sustainable use of water resources could be implemented through the establishment of a reasonable management framework [3,4]. From the perspective of global water use Wada et al. (2014) believed that the main ways could be explored and to improve the sustainability of water resources through the reconstruction of the past and the prediction of the future [5]. From the point of view of European water usage, water footprint methods were applied to evaluate the sustainability of water resources by Vanham (2017) [6]. From the perspective of promoting water management in watersheds, the assessment methods to promote regional sustainability by improving the sustainability of watershed water management has been studied (Cortes, 2012) [7]. Singh (2020), Fard (2020), from the perspective of urban water management in developing countries, discussed the of water resources sustainability through case studies of cities in India and Iran [8,9].

The sustainable management and evaluation methods of water resources has been also explored by domestic scholars from different levels. LiShuang (2019), LiYunjie (2016), GuoLidan (2020) and ZhangQian (2019) analyzed and evaluated the sustainable use of water resources in watersheds, provinces and cities using water ecological footprint method [10-13]. SunYaru (2019) and Liu Li (2020) used cloud models to analyze and evaluate the water safety of Nanjing city and the water sustainability level and spatial heterogeneity of nine cities in the PearlRiver delta respectively [14,15]. ZuoQiting (2020) used TOPSIS method to analyze the of water resources carrying capacity in nine provinces of the Yellow River Basin [16], and AiYadi (2020) used AHP-entropy weight method to evaluate the degree of water resources development in Xi'an [17]. For the northern part of China, water comes from underground, unlike surface water, groundwater has its own characteristics. For this type of region, scholars also used different analytical methods to study. Dong Yi (2019) selected the Bossel index system to evaluate the underground in Lishu County, Jilin Province [18]. Wang Lei (2019) used the AHP method to evaluate the groundwater safety in Yanbian Prefecture and put forward the countermeasures and for sustainable utilization [19].

For the evaluation of sustainable use of water resources, scholars at home and abroad mainly focus on scarcity, safety and carrying capacity, but from a systematic point of view, the sustainability of water resources has been studied by time axis, and the research is still in frame stage [20], especially the lack of practical evaluation. In this paper, water resources are regarded as a system, and the evaluation model of sustainable utilization of water resources has been established, and Wuhan has been taken as an example to make an empirical analysis.

### **3. Overview of the Study Area**

Taking Wuhan as an example, this paper analyzes the sustainability of urban water system. Wuhan is the capital city of Hubei Province, known as the province of thousands of lakes, and the central city of central China.

Wuhan is located at the junction of the Yangtze River and the Han River and the eastern part of the Jianghan Plain. The landform belongs to the transition area of southeast Hubei hilly area from the eastern margin of the Han River Plain to the southern foot of Dabie Mountain. The north subtropical monsoon climate has the

following characteristics, perennial abundant rainfall, abundant heat, cold winter and hot summer, distinct four seasons. There are many lakes in the city, accounting for about one quarter of the city's land area and a surface area of 2117.6km<sup>2</sup>. There are 165 rivers within the urban length above 5km, the surface area of 471.31km<sup>2</sup>; 166 lakes listed in the protection list; the control area of the city is 867.07km<sup>2</sup>; 264 reservoirs with a total storage capacity of 875 million m<sup>3</sup>.

With the increase of population, the acceleration of urbanization and the need of economic development, the water demand of cities increases greatly. At the same time, due to the instability of precipitation and the problems of surface water quality, there are still some problems in the supply, use and reuse of urban water system.

#### **4. Data sources and Research Methods**

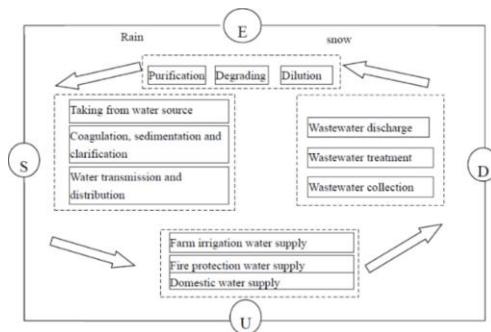
#### *4.1. Data Sources*

The data in this paper are from China Statistical Yearbook [21], Hubei Statistical Yearbook [22], Wuhan Statistical Yearbook [23] and Wuhan Water Resources Bulletin [24].

#### *4.2. Research Methods*

#### *4.2.1. Determination of Evaluation Models*

SUDE model is a sustainable model based on the whole life cycle of water resources. The analysis of the model regards water medium as a natural resource, which is used by human society to meet the needs of normal residents, industrial production, fire protection and farmland irrigation, and the water after use is collected and treated by drainage pipe system according to the different degree of pollution. After reaching the drainage water quality index, this part of the water goes back to nature--into rivers, lakes and seas, or recharge groundwater. Therefore, water resources can be divided into four stages of life cycle, that is, resource supply stage ("S"), water use stage ("U"), water collection and treatment stage ("D") and water return to natural water ecological ("E") environment for dilution and degradation. SUDE model is established, as shown in figure 1.



**Figure1.**The model of SUDE.

#### 4.2.2 Determination of the Evaluation Method

The entropy weight TOPSIS method is the synthesis of entropy weight and TOPSIS method. Firstly, the method uses the amount of information carried by the evaluation index data, that is, the entropy weight method is used to give the index weight, and then the evaluation object is calculated and sorted by the TOPSIS method. On this basis, the optimal evaluation object is obtained. This evaluation method belongs to the objective evaluation method, which can make full use of the information of the data and effectively eliminate the influence between different indexes by the same trend and normalization of the original data.

The paper firstly assimilates and normalizes the original data to eliminate the influence between different indexes, then calculates the weight of each evaluation index in the SUDE model by entropy weight method, and then uses the TOPSIS to calculate the ranking of the evaluation objects.

### 5. Establishment of the Evaluation System and the Index Calculation

#### 5.1. Establishment of the Evaluation System

Based on the above analysis, we can establish the evaluation indicators of SUDE, as shown in Table 1.

**Table 1.** Water system sustainability evaluation index system based on SUDE.

three-level indicators					weight of indicators
the first	the Second	the third		the Second	the third
S	S1	Total volume of water resources ( $10^8m^3$ )		+	0.0927
	S2	Total volume of water supply ( $10^8m^3$ )		-	0.0430
	S3	Transit Water volume ( $10^8m^3$ )		+	0.0275
	S4	The percentage of water quality standards in the first-level water functional area (%)		+	0.2434
	S5	Length of the water supply pipe network(km)		+	0.0383
	U1	Water consumption volume of ten thousand yuan GDP ( $10^8m^3$ )		-	0.0419
U	U2	Water consumption volume of farmland irrigation per mu area ( $m^3$ )		-	0.0307
	U3	Water consumption volume of the industrial added value of ten thousand yuan ( $m^3$ )		-	0.0651
	U4	Domestic water consumption volume of urban residents per person per day (/)		-	0.0273
	U5	Domestic water consumption volume of urban residents per person per day (/)		-	0.1968
	D1	The percentage of water supply population in central areas (%)		+	0.0396
	D2	The average volume sewage discharge per person per year ( $10^4m^3$ )		-	0.0341
D	D3	Total volume of sewage discharge of industrial wastewater ( $10^4m^3$ )		-	0.0281
	D4	Total volume of COD discharge in the industrial wastewater (m $3$ )		-	0.0330
	D5	Total volume of Ammonia nitrogen discharge in the industrial wastewater (m $3$ )		-	0.0623
	D6	Total volume of COD discharge of urban domestic sewage (m $3$ )		-	0.0542
	D7	Total volume of Ammonia nitrogen discharge in the urban domestic sewage (m $3$ )		-	0.0515
	D8	The percentage of industrial water reuse (%)		-	0.0334
E	D9	The percentage of centralized treatment of urban domestic sewage (%)		+	0.0312
	D10	Length of the drainage pipe (km)		+	0.0617
	E1	Investment in ecological construction and protection completed within the year (10 $4$ yuan)		+	0.0277
	E2	Investment in ecological construction and protection completed within the year (10 $4$ yuan)		+	0.0312
	E3	The average area of park green space occupied by per person (10 $4$ ha)		+	0.0227
	E4	Greening coverage rate in built-up areas		+	0.0266
E	E5	Length of the levee		+	0.0413
	E5	Direct economic loss value of geological disasters		-	0.0218

## 5.2. The Index Calculation

The steps for the calculation are as follows.

(1) Establish a matrix composed of the original index data

Establish an original index data matrix that indicates the number of samples as "m", evaluation index is "n", as follows:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \quad (1)$$

(2) Assimilation and normalization processing of data

1) The positive and negative Indicators are assimilated to obtain the matrix  $X'_{ij}$

$$\text{The positive Indicators: } x'_{ij} = \frac{(x_{ij} - x_{\min})}{(x_{\max} - x_{\min})} \quad (2)$$

$$\text{The negative Indicators: } x'_{ij} = \frac{(x_{\max} - x_{ij})}{(x_{\max} - x_{\min})} \quad (3)$$

In the formula above,  $x_{\max}$  is the maximum of the j indicator and  $x_{\min}$  is the minimum of the j indicator.

2) The data matrix is normalized to obtain a standardized matrix consisting of processed data, namely

$$Z_{ij} = \frac{x_{ij}}{\sqrt{\sum(x_{ij})^2}} \quad (4)$$

$$Z = \begin{bmatrix} z_{11} & \cdots & z_{1n} \\ \vdots & \ddots & \vdots \\ z_{m1} & \cdots & z_{mn} \end{bmatrix} \quad (5)$$

(3) Determination of the weights of the index data in the matrix

The weight of the index data is determined using the entropy weight method.

1) After the normalization of the matrix  $X'_{ij}$ , a dimensionless treatment is performed to obtain the matrix  $P_{ij}$

$$P_{ij} = \frac{x'_{ij}}{\sum_{i=1}^m x'_{ij}} \quad (6)$$

2)

$$e_j = \frac{-1}{\ln m} \sum_{i=1}^m P_{ij} \times \ln P_{ij} \quad (7)$$

$$d_j = 1 - e_j \quad (8)$$

$$W_j = \frac{d_j}{\sum_{j=1}^n d_j} \quad (9)$$

$$3) Z_{ij}^* = W_j \times Z_{ij} \quad (10)$$

(4) Calculating the optimal solution  $Z_j^{*+}$  And the worst solution  $Z_j^{*-}$

(5) Calculating the distance between the evaluation object and the optimal solution  $D_i^+$ ; calculating the distance between the evaluation object and the worst solution  $D_i^-$ , and then determining the comprehensive index  $C_i$

$$D_i^+ = \sqrt{\sum_{j=1}^n (Z_{ij}^{*+} - Z_{ij})^2} \quad (11)$$

$$D_i^- = \sqrt{\sum_{j=1}^n (Z_{ij}^{*-} - Z_{ij})^2} \quad (12)$$

$$C_i = \frac{D_i^-}{D_i^+ + D_i^-} \quad (13)$$

Computing the value of  $C_i$ , the comprehensive index of each evaluation object. And judge the order of the scheme according to the size of its value. Then, combined with the evaluation level standard (As shown in Table 2), the evaluation object, namely, the level of the sustainable utilization of the water resources can be obtained. We also could find the change of water utilization of Wuhan water resources (as shown in figure 2).

**Table 2.** The evaluation level standard of the sustainable utilization of the water resources.

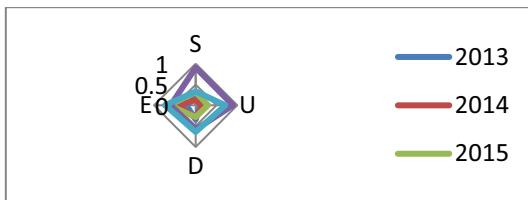
closeness	[0,0.2]	(0.2,0.4]	(0.4,0.6]	(0.6,0.8]	(0.8,1.0]
the evaluation level standard	Weak	Relatively weak	Medium level	Relatively strong	strong
	V	IV	III	II	I

## 6. Analysis of the Evaluation Results of the Sustainable Utilization of Water Resources

According to the sample data of 2013~2017, the sustainable utilization level of Wuhan water system is calculated. The calculation results including the weight of indicators and the level of the sustainable utilization are shown in Table 1 and 3 respectively. According to the calculation results, the sustainable utilization level of the water system in Wuhan is showing a good development trend, and the sustainable utilization capacity is gradually improved. Among them, the water system sustainable utilization comprehensive indexes of 2013 and 2014 were 0.3194 and 0.3455, respectively, which were at the lower level. While the subsequent index of 2015 year reached 0.4319, rising to the medium level. The indexes of 2016 and 2017 were divided into 0.6657 and 0.6147, respectively, showing that the sustainable utilization of water resources in Wuhan has entered a good level of development.

**Table 3.** The level of the sustainable utilization of Wuhan water resources (2013~2017).

	$d_i^+$	$d_i^-$	$c_i$	level
2013	0.1583	0.0574	0.2660	IV
2014	0.1515	0.0482	0.2414	IV
2015	0.1273	0.0618	0.3332	III
2016	0.0613	0.1492	0.7088	II
2017	0.1083	0.1220	0.5296	III

**Figure2.** the Change of water utilization of Wuhan water resources (2013~2017).

(1) Analysis the four subsystems of the SUDE (As shown in Table 4), the development index of the subsystem S ranged from the lower level of 2013~2015 years, continuously increasing thereafter, especially in 2016 year, reaching the peak. The main reasons of the index increasing depends on the increasing of the total volume of water resources and the transit water volume. That year had 29.4% more rainfall than in the past year and the volume of rainfall in that year was 46.2% more than the multi-year average. In 2017 year, the volume of rainfall year was 6.7% less than the annual average, and water supply levels declined at the sametime.

**Table 4.** Sustainable utilization index value of SUDE system and its subsystem of Wuhan (2013~2017).

	S	U	D	E
2013	0.2088	0.1072	0.3445	0.2229
2014	0.1269	0.1660	0.2958	0.4438
2015	0.3199	0.2967	0.2827	0.6340
2016	0.9077	0.8975	0.5815	0.5999
2017	0.3344	0.7109	0.6393	0.7307

(2) The value of development index of the subsystem U of SUDE was in the [0,0.2] interval with a low level of sustainable development in 2013 year, and was in the interval of (0.2,0.4] with a relatively low level of in 2014 and 2015 year. After 2016, it showed a highlevel of development trend. Relatively speaking, the level of urbanization in 2013~2015 was not very high, so the demand for domestic water consumption volume of urban residents per person per day was relatively low. But at the same time, due to the lack of science and technology level and the lack of water-saving consciousness, the number of agricultural mu average irrigation water and industrial 10,000 yuan added value water in this period remained high. But at the same time, due to the low level of science and technology and the lack of water-saving consciousness, both Water consumption volume of farmland irrigation per mu area and Water consumption volume of the industrial added value of ten thousand yuanin this period remained high.

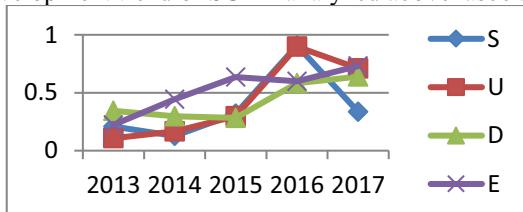
Especially after 2016, the proportion of water supply population in the central urban area increased, both the sanitary conditions of the whole society and the water supply efficiency of the water system were improved, which ensured that the relative levels of the subsystem U in both 2016 and 2017 year were higher.

(3) In the subsequent system D, the discharge volume of industrial wastewater and urban domestic sewage, COD and ammonia nitrogen emissions, although repeated but

showed an overall downward trend, indicating the improvement of drainage water quality, which is quite friendly to the environmental system. However, as the important indexes, which were reflected in the industrial and domestic sewage treatment capacity—the percentage of industrial water reuse and the percentage of centralized treatment of urban domestic sewage, were continuously improved, which depends on the increase of Wuhan municipal government investment in the environmental foundation such as drainage pipe network and the enhancement of the supervision of the sewage treatment level of industrial enterprises and sewage treatment plants.

(4) Similar to the system D, the sustainable level of subsystem E increased year by year after experiencing lower levels in 2013. The investment of ecological construction and protection by Wuhan Municipal Government ensured the improvement and development of ecological aspects such as the average area of park green space occupied by per person, greening coverage rate in built-up areas and the length of the levee. And at the same time, the sustainability of the whole ecological subsystem E has been also ensured.

The sustainable development trend of SUDE analyzed above has been shown in figure 3.



**Figure 3.** Sustainable development trend of SUDE subsystem in Wuhan (2013~2017).

## 7. Conclusion and Suggestions

(1) Compared with the previous evaluation index system, the urban water system based on SUDE model can analyze the problems from the system perspective, help to identify and select some important relevant factors which belong not only to the aspects of technology but also to the aspects of technology management, and better reflect the internal connection between the elements of the urban water system. In addition, the selection of indicators also considers the operability and reliability of data collection.

(2) Efforts should be strengthened continuously to ensure the supply quality of water resources. In particular, we should not rely on the endowment of water resources, efforts should be made to improve the compliance rate of the first-level water functional zones and strengthen the construction of the water supply pipe network.

(3) Improve the efficiency of water use, increase the promotion of agricultural irrigation technology, and continuously strengthen water saving publicity among urban residents, improve the awareness of water saving, while ensuring that the quality of living standards not decline. The total amount of water would be appropriately reduced, so as to reduce the pressure of urban water supply.

(4) Continue to strengthen supervision, especially for the supervision of drainage points near surface water, strengthen enterprise regular inspection and sampling inspection, so as to ensure the normal operation of sewage treatment facilities, which can not only improve the circulation rate of water treatment to reduce the volume of sewage water, but also reduce the water quality of sewage water, to achieve the dual control goal of quality and quantity.

(5) Increase investment in ecological construction. While nature gives us water resources through rainfall, it also threatens us with possible disasters such as floods. With preventing disasters in the bud, increasing ecological protective efforts, increasing the protection and construction of green space and wetlands, improving the runoff coefficient of drainage to improve the natural water storage capacity, we could relieve the transportation and discharge pressure of the artificial drainage system, so as to ensure that the city passes safely during the once-in-a-century flood disaster.

## References

- [1] Niu W. Introduction to sustainable development. Sci. Press; 1994.1-123 p.
- [2] Feng S. Introduction to the continuous utilization and management of water resources. Sci. Press; 2000.62-214 p.
- [3] Koop S, Leeuwen CJ. Assessment of the sustainability of water resources management: a critical review of the city blueprint approach. Water Resources. 2015 Sept; 29(15):5649-70.
- [4] Kasim A, Gursoy D, Okumus F, Wong A. The importance of water management in hotels: a framework for sustainability through innovation. J. of Sustainable Tourism. 2014 Jul;22(7):1090-107.
- [5] Wada Y, Bierkens MF. Sustainability of global water use: past reconstruction and future projections. Environ. Res. Lett. 2014 Oct; 9 (10):1-17.
- [6] Vanham D. The water footprint of the EU: quantification, sustainability and relevance. Water Int. 2018.Jun; 43(6):731-45.
- [7] Cortés AE, Oyarzún R, Kretschmer N, Chaves H, Soto G, Soto M, Amezaga J, Oyarzún J, Rotting T, Senoret M. Application of the watershed sustainability index to the elqui river basin north-central chile. Obras Y Proyectos. 2012.12:57-69.
- [8] Singh AP, Bhakar P. Development of groundwater sustainability index: a case study of western arid region of Rajasthan, India. Environment Development and Sustainability. 2020.2.
- [9] Fard ZD, Rahimi M, Javid N. Studying radon concentration in drinking water resources in Zarand city (Iran) and its villages. J.R.N..2020 Aug; 326 (8):33-9.
- [10] Li S, Du JK, Xing HH, Wang SX, Li F. Sustainable utilization evaluation of water resources in the main stream of Hangjiang river based on water footprint theory and grey target model. Water Saving Irrigation. 2019 Sept; (9):74-80.
- [11] Li YJ, Lu HJ, Bu P. Sustainable utilization of water resources in Zhejiang province based on ecological footprint analysis. J. of Yangtze River Scientific Research Institute. 2016 Dec;33(12):22-6,32.
- [12] Guo LD, Jing PR. Evaluation of sustainable utilization of water resources in Jiangsu province based on ecological footprint theory. J. of Economics of Water Resources. 2020 Mar; 38(3):19-26.
- [13] Zhang Q, Xie SY. Using Ecologicalwater footprint model to analyze sustainable use ofwater resources in Chongqing.J. of Irrigation and Drainage. 2019 Mar; 38(3):93-100.
- [14] Sun YR, Dong ZC, Xv Y, Bao QY. Urban water security evaluation based on cloud model. Yellow River. 2019 Aug; 41(8):52-6.
- [15] Liu L, Wang L. The entropy weights-based approach to and the normal cloud model for water resources assessment.J. of South China Normal University(Natural Science Edition). 2020 Jan;52(1): 77-84.
- [16] Zuo QT, Zhang ZZ, Wu BB. Evaluation of water resources carrying capacity of nine provinces in Yellow River basin based on combined weight topsis model. Water Resources Protection. 2019 Feb;36(2):1-7.
- [17] Ai YD, Wei CJ, Ma ZZ. Evaluation on water resources development and utilization degree based on ahp-entropy weight method. Advances in Science and Technology of Water Resources. 2020 Feb;40(2):11-6.
- [18] Dong Y, Liang XJ, Xiao CL, Wang YF. Sustainable utilization assessment of groundwater resources in Lishu county based on bossel index system. Water Power. 2019 Jan;45(1): 9-12.
- [19] Wang L, Liang XJ, Li HW, Lu H. Study on sustainable utilization of groundwater resources in Yanbian prefecture based on water safety. Yangtze River. 2019 May;50(5):94-8.
- [20] Gao CB, Zeng HY, Chen XG, Chen JW. Evaluation of environmental sustainability for urban water systems based on lca framework. Water Resources Protection. 2007 Feb;23(2):51-3.
- [21] National Bureau of Statistics of China. China statistical yearbook 2013-2018. China Statistics

- Press;2019.
- [22] National Bureau of Statistics of China. Hubei statistical yearbook 2013-2018. Beijing: China Statistics Press; 2019.
- [23] Wuhan Minicipal Statistics Bureau. Wuhan statistical yearbook 2013-2018. Beijing: China Statistics Press; 2019.
- [24] Wuhan Municipal Water Resources Bureau.Wuhan Water Resources Bulletin<http://swj.wuhan.gov.cn/szy>

# An Empirical Study on the Relationship Between Energy Consumption and GDP in China

Jian ZHANG<sup>a</sup>, Xin ZHOU<sup>b,c1</sup>

<sup>a</sup>*School of Economics and Management, Shanghai University of Political Science and Law, Shanghai, 201701, China*

<sup>b</sup>*School of Economics, Shanghai University, Shanghai, 200444, China*

<sup>c</sup>*School of Languages and Cultures, Shanghai University of Political Science and Law, Shanghai, 201701, China*

**Abstract.** Based on the provincial panel data of China from 1990 to 2018, this paper establishes a simultaneous equation model containing exogenous variables such as capital stock, urbanization and trade liberalization to explore the causal relationship between energy consumption and economic growth. The study found that China's GDP and energy consumption influence each other, which is manifested as an endogenous causal relationship in the economic system. When the economic aggregate increases, the demand for energy has risen sharply. However, due to the low energy use efficiency production mode and environmental destruction and pollution, energy consumption shows a negative scale effect that restricts economic growth. China's economic growth is manifested in diminishing returns to scale, and the urbanization rate and trade openness have a significant impact on energy consumption.

**Keywords:** Energy consumption, economic growth, panel data simultaneous equations model

## 1. Introduction

Studying the relationship between energy consumption and GDP has important application value for a country's policy makers. On the one hand, the government is highly concerned about the speed of economic growth. On the other hand, the government is worried about the destruction of the environment and the excessive consumption or lack of resources. If energy consumption does not have a significant impact on GDP, the government can more easily formulate policies to protect the environment and resources without worrying about the decline in economic growth. If there is a significant relationship between energy consumption and GDP, then the policies for protecting the environment and resources will limit economic growth and hinder the improvement of residents' welfare, which makes the implementation of energy saving policies very difficult.

---

<sup>1</sup>Corresponding Author: Xin Zhou, School of Economics, Shanghai University, Shanghai, 200444, China; Email: zhouxin@shupl.edu.cn.

The relationship between energy consumption and GDP has not only attracted the attention of policy makers, but also attracted the attention of many scholars. They used the data from different countries and regions in combination with various econometric methods to conduct research of the two variables, and obtained conclusions with huge differences. At present, the academic circles believe that there are four situations in the relationship between energy consumption and GDP. The first is the hypothesis of environmental protection, which believes that changes in GDP are the cause of changes in energy consumption. Therefore, any policy to protect the environment and resources will not affect the economy growth. The second is the growth hypothesis, which believes that energy consumption is the cause of GDP changes. Therefore, any policy to reduce energy consumption will limit the growth of GDP. The third is the feedback hypothesis, which believes that there is a two-way causal relationship between energy consumption and GDP. The fourth is the irrelevant hypothesis, which believes that there is no causal relationship between energy consumption and GDP.

Since previous studies have not obtained consistent and reliable conclusions, it is difficult for policy makers to decide which energy policies to adopt for economic development. Most previous empirical studies were based on single equation models that lacked economic theoretical support. The studies mainly examining the causal relationship in a single direction from independent variables to dependant variable, and lacked the discussion of the mechanism of simultaneous interaction between variables. At the same time, they used time series data for analysis. The changes in economic variables in reality often show non-stationary and nonlinear characteristics, such as production functions, etc. Therefore, the robustness of the research conclusions has been greatly questioned. Based on China's provincial panel data from 1990 to 2018, we use the panel simultaneous equation model to explore whether there is a causal relationship between energy consumption and economic growth within a endogenous system, so as to accurately formulate macroeconomic development policies to promote economic growth and improve energy use efficiency.

## **2. Literature Review**

The research on the relationship between energy consumption and economic growth can be traced back to the 1970s. In Kraft & Kraft's pioneering study in 1978, they found that GNP was the cause of changes in energy consumption by using US data from 1947 to 1974.[1] Subsequently, other scholars used time series data from different countries to conduct causal analysis on the relationship between economic growth and energy consumption (e.g., Akarca and Long 1979, [2] Akarca and Long 1980, [3] Eden and Hwang 1984 [4], Yu and Choi 1985, [5] Erol and Eden 1987 [6]). They found that the direction of influence between economic growth and energy consumption is not fixed and the research conclusions are different. In most cases, studies on the relationship between economic growth and energy consumption use ordinary least squares to estimate log-linear models and perform statistical tests on coefficients. The defect is that the non-stationary characteristics of time series data are not considered in previous studies. Therefore, their conclusion may be wrong.

In the past few decades, time series analysis methods have made rapid development and progress. During the same period, the introduction of panel data models has also improved the robustness of the estimation results. New methods beyond simple linear regression are used to study the relationship between energy consumption and economic

growth. The researches of correlation in this period were mostly based on the Engle-Granger test, Johansen cointegration test, vector error correction model and vector autoregressive model (e.g., Eden and Jin 1992 [7], Masih and Masih 1996 [8], Masih and Masih 1997 [9], Glasure and Lee 1997 [10], Asufu-Adjaye 2000 [11], Hondroyannis etc. 2002 [12], Soytas and Sari 2003 [13], Altinay and Karagol 2004 [14], Ang 2007 [15], Ang 2009 [16], Halicioglu 2009 [17], Jalil and Mahmud 2009 [18], Apergis and Payne 2009 [19], Apergis and Payne 2010 [20], Hossain 2011 [21]). By analysing the time series data and panel data of different countries including China, the conclusions are still inconsistent. The main reason for the different research results in the above is mainly due to the different econometric methods used in the empirical analysis.

By reviewing the relevant research above, it can be found that testing the same indicators in the same country could lead to different conclusions. The reasons might be different periods of research and different methods used. Another common reason is insufficient sample data for the study. The time span of these studies using time series models is 30 to 40 years. Small sample size may lead to lower reliability of statistical tests, so that the results of the research are not consistent. In order to solve the statistical defects of the time series model due to the lack of sample size, the panel data method is applied to the study of the relationship between energy consumption and economic growth. However, the application of panel data methods has also brought new problems. Regarding different countries as a homogeneous group rather than independent individuals cannot identify the different dynamic relationships between energy consumption and economic growth. When other explanatory variables are added in the empirical models, the relationship between energy consumption and economic growth could also change due to the different levels of economic development among countries.

This paper uses China's provincial panel data to establish a simultaneous equation model to study the relationship between GDP and energy consumption. The system considers capital stock, urbanization rate, square of GDP, and trade liberalization rate as exogenous variables. It could not only improve the statistical defects caused by the lack of sample size in the time series model, but also avoids the problem of heterogeneous individuals in panel data model. It also depicts the dynamic relationship between different variables while constructing models based on economic theory and improving the statistical reliability of research results.

### 3. Data and Model

#### 3.1. Selection of Data and Variable

We select data of per capita GDP, per capita energy consumption, per capita capital stock, urbanization rate, and trade openness of 30 provinces, cities, and autonomous regions in China from 1990 to 2018 (Because Tibet lacks energy consumption data, the panel data simultaneous equation model only contains data for 30 provincial units.), and establishes a panel simultaneous equation model to study the relationship between economic growth and energy consumption. Each variable uses the GDP deflator as a measure of price index to adjust the relevant variables at constant prices. In order to maintain the consistency and timeliness of the data, 2018 is selected as the base period. Among these variables, the capital stock of each province is calculated by the perpetual inventory method, the urbanization rate is calculated by the ratio of the urban population

to the total population, and the trade openness is calculated by the ratio of the total import and export trade to GDP.

### 3.2. Data Stationarity Test

In order to construct a simultaneous equation model, we examine the stationarity of the variables by performing a unit root test at first. Since the time series data of per capita GDP, per capita energy consumption, and per capita capital stock show a strong trend, we take the logarithm of the three variables for further study. While the urbanization rate and trade openness are proportional data which no trend were shown, so their original forms were used for modeling. The PP-Fisher panel unit root test results of all five variables are shown in Table 1. *lgdp* represents the logarithm of per capita GDP, *lec* represents the logarithm of per capita energy consumption, and *lcap* represents logarithm of the per capita capital stock, *to* represents trade openness, *urb* represents urbanization rate. The test statistics of the level values of each variable in Table 1 are all less than the critical value of the 5% significance level, indicating that the variables are not stationary. The test statistics of the variables after the first order difference are all greater than the critical value of the 5% significance level. The test results indicate that all variables are stable after first order difference.

**Table 1.** The results of PP-Fisher panel unit root test.

	<i>lgdp</i> Statistic	<i>lec</i> Statistic	<i>lcap</i> Statistic	<i>to</i> Statistic	<i>urb</i> Statistic
<b>Level</b>	28.5700 <sup>a</sup>	40.8100	37.1100	68.1500	70.5700
<b>1st Difference</b>	157.2300	214.3400	117.2400	567.5000	304.4200

<sup>a</sup> The Chi-square critical value of the 5% significance level is 79.0819.

### 3.3. Panel Cointegration Test

In order to analyse the relationship between economic growth and energy consumption in a model system, it is necessary to determine whether there is a co-integration relationship between economic growth and energy consumption while per capita capital stock, trade openness and urbanization rate are taken as exogenous variables. Per capita GDP and per capita energy consumption are tested by panel co-integration. There is a cointegration relationship between *lgdp* and *lec* according to the statistical *p* value of the trace test and the maximum eigenvalue test results in Table 2 under the significance level of 5%. The simultaneous equation model can be further established.

**Table 2.** The results of Johansen co-integration test.

	Trace Statistic	Trace Prob.	Max-Eigen Statistic	Max-eigen Prob.
<b>None</b>	25.8700	0.0010	23.8900	0.0010
<b>At Most 1</b>	1.9800	0.1590	1.9800	0.1590

### 3.4. Panel Granger Causality Test

In order to construct a simultaneous equation model of economic growth and energy consumption, it is also necessary to examine whether there is a two-way causal relationship between per capita GDP and per capita energy consumption. If there is no two-way causal relationship between the two variables, then the simultaneous equation model cannot be constructed. Therefore, the panel data Granger test between *lgdp* and *lec* is further carried out. The statistical *p* values of the test results in Table 3 are all less than the 5% significance level. It means that there exists a two-way causal relationship

between per capita GDP and per capita energy consumption, and the simultaneous equation model can be further established.

**Table 3.** The results of panel Granger causality test.

Null Hypothesis	F-Statistic	Prob.
Lec does not Granger Cause lgdp	12.4600	0.0000
lgdp does not Granger Cause lec	3.5100	0.0300

### 3.5. Construction and Estimation of Simultaneous Equation Model

According to the results of various tests before, considering to establish a simultaneous equation model with a total production function equation and a Kuznets energy consumption function equation. We propose an aggregate production function model that includes the spillover effect of energy consumption on productivity. It is assumed that energy consumption is an important input factor in production, and the aggregate production function of each province is set as follows:

$$GDP_{it} = CAP_{it}^{\beta} (A_0 POP_{it})^{1-\beta} EC_{it}^{\gamma} \quad (1)$$

Where  $GDP_{it}$  is the real GDP of province  $i$  at time  $t$ ,  $CAP_{it}$  is the real capital stock of province  $i$  at time  $t$ ,  $POP_{it}$  is the population of province  $i$  at time  $t$ , and  $EC_{it}$  is the per capita energy consumption of province  $i$  at time  $t$ ,  $A_0$  is the initial technical level,  $\beta$  ( $0 < \beta < 1$ ) and  $\gamma$  ( $-1 < \gamma < 1$ ) are parameters to be estimated. If  $\gamma=0$ , it means that energy consumption has no spillover effect on economic growth, and the aggregate production function is constant returns to scale. If  $\gamma<0$ , it means that economic growth is manifested as diminishing returns to scale, and energy consumption has a reverse spillover effect on economic growth. It means that the use of energy under the current production mode limits further economic development. The inefficient use of energy and the limited petrochemical resources dragged down the economic growth, and reduce the impact of other production factors on output. If  $\gamma>0$ , it means that the aggregate production function shows as increasing returns to scale. Energy has a positive spillover effect on economic growth. The increase in the use of energy will promote the increase in output. After taking the logarithm of all variables and adding a random disturbance term, equation (1) can be rewritten as follows:

$$\lgdp_{it} = c + \beta \cdot \lgcap_{it} + \gamma \cdot \lgc_{it} + u_{it} \quad (2)$$

Where  $\lgdp_{it}=\ln(GDP_{it}/POP_{it})$  is the logarithm per capita GDP of province  $i$  at time  $t$ ,  $\lgcap_{it}=\ln(CAP_{it}/POP_{it})$  is the logarithm per capita capital stock of province  $i$  at time  $t$ ,  $\lgc_{it}=\ln(EC_{it})$  is the logarithm per capita energy consumption of province  $i$  at time  $t$ ,  $c=\ln(A_0)$  is the intercept term which means the initial technical level,  $u_{it}$  is the random disturbance term.

Chinese government advocates the Belt and Road Initiatives to promote international trade among the countries. They also propose urbanization within China to accelerate the transfer of rural population to urban to promote economic development and improve the living standards of residents. So that the impact of trade openness and urbanization on energy consumption and economic growth cannot be ignored. Therefore,

a corresponding environmental Kuznets curve equation including urbanization rate and trade openness is established for energy consumption:

$$lec_{it} = \beta_0 + \beta_1 lgdp_{it} + \beta_2 lgdp_{it}^2 + \beta_3 urb_{it} + \beta_4 to_{it} + \varepsilon_{it} \quad (3)$$

Where  $lec_{it}$  is the logarithm per capita energy consumption of province  $i$  at time  $t$ ,  $lgdp_{it}$  is the logarithm per capita GDP of province  $i$  at time  $t$ ,  $urb_{it}$  is the urbanization rate of province  $i$  at time  $t$ ,  $to_{it}$  is the trade openness of province  $i$  at time  $t$ ,  $\varepsilon_{it}$  is the random disturbance item. The signs of  $\beta_1$ ,  $\beta_2$  and  $\beta_4$  are assumed positive, indicating that the increase in GDP, urbanization rate, and trade openness will increase energy consumption. If the environmental Kuznets curve hypothesis is valid in China, the sign of  $\beta_2$  should be negative, which means that after the economy develops to a certain extent, the total energy consumption begins to decline, and the economic growth mode changes from extensive to intensive.

The econometric methods of estimating parameters used in previous studies such as ordinary least squares method, instrumental variable method, and maximum likelihood method have their own limitations. Their parameter estimators have good statistical properties only when the model meets certain assumptions. For example, the parameter estimator of the maximum likelihood method is a reliable estimator only when the error term of the model obeys a normal distribution or a certain known distribution. The Generalized Method of Moments (GMM) does not require special assumptions of error term. It allows error terms to have heteroscedasticity and serial correlation, and does not need to know the exact distribution of the error term. The obtained parameter estimates from GMM are more realistic and robust than other parameter estimation methods. Therefore, we use GMM method to estimate the panel simultaneous equation model. The results are shown in Table 4.

**Table 4.** The results of simultaneous equation GMM-HAC method.

	The Results of Equation (2)		The Results of Equation (3)
$lec_{it}$	-0.1584*** <sup>a</sup> (0.0529) <sup>b</sup>	$lgdp_{it}$	2.0813*** (0.4275)
$lcap_{it}$	0.4938*** (0.0639)	$lgdp_{it}^2$	-0.0623*** (0.0203)
		$urb_{it}$	0.0020* (0.0012)
		$to_{it}$	0.0003* (0.0002)
Intercept	8.0415*** (1.2528)	Intercept	-7.9865*** (2.3960)
$Adj. R^2$	0.9992	$Adj. R^2$	0.9923
DW	1.8451	DW	2.0105

<sup>a</sup> \*\*\* and \* mean that the corresponding coefficients pass the significance test at the significance level of 1% and 10% respectively.

<sup>b</sup> The standard deviation of the corresponding estimator is in parentheses.

The estimation result of equation (2) shows that per capita capital stock and per capita energy consumption have a significant impact on per capita GDP. Every 1% increase in per capita capital stock will increase per capita GDP by 0.4940%. This also shows that the output elasticity of the capital stock is basically equal to the output elasticity of the population in China. The coefficient of per capita energy consumption

$\gamma = -0.1580$  is less than zero and significant at 1% significance level. It means that in the current economic development mode, energy consumption has a reverse spillover effect on GDP, and economic growth affected by energy use efficiency, it is manifested as diminishing returns to scale.

The estimated result of  $\beta_1$  in equation (3) shows that the economic growth mode is intensive. If per capita GDP increases by 1%, per capita energy consumption will increase by 2.0813%. In the long run, per capita energy consumption will begin to decline with the growth of per capita GDP only when the per capita GDP exceed 17,963,497 ¥. It means that in accordance with China's current industrial structure, energy policy, and production mode, it is impossible to reduce energy consumption by simply relying on the improvement of economic development.

The coefficient of the urbanization rate is positive and significant at 10% significance level. It means that the increase in urbanization rate will increase the total energy consumption. If the urbanization rate increases by 1 percentage, per capita energy consumption will increase by 0.2000%. The coefficient of trade openness is positive and significant at 10% significance level. If the trade openness increases by 1 percentage, per capita energy consumption will increase by 0.0300%. The results mean that although China's economic growth is largely dependent on import and export trade, but the impact of trade on energy consumption is relatively weak.

## 4. Conclusion

China has successively proposed the Belt and Road Initiative, promoted urbanization and implemented supply-side structural reforms. Since China is a traditional energy demand country, the relationship between energy consumption and economic growth has attracted more and more attention from academic circles and Chinese government decision-making departments. This paper establishes a simultaneous equation model based on growth theory to empirically study the relationship between energy consumption, GDP, capital stock, urbanization and trade openness in China with the use of panel data of 30 provincial administrative regions from 1990 to 2018. The estimation results of simultaneous equation show that:

- 1) There is a long-term stable two-way causal relationship between energy consumption and GDP.
- 2) The environmental Kuznets curve hypothesis is valid in China.
- 3) The energy consumption has a reverse spillover effect on GDP.
- 4) The economic growth expressed as a production function is diminishing returns to scale.
- 5) The Increase in urbanization leads to a slight risen in energy consumption.
- 6) The increase in trade openness will increase energy consumption, but the impact is relatively weak.

It is concluded from the empirical analysis results that China's GDP and energy consumption are endogenous causal relationship within the economic system under the condition of controlling capital stock, urbanization rate and trade openness. The study also finds that the growth of energy consumption shows a negative influence on economic development under the condition of scarce resources, the extensive production mode, environmental damage and pollution. China's inefficient energy use leading to diminishing returns to scale in economic growth. The above results not only conform to

the theoretical relationship between economic growth and energy consumption, but also conform to the current status and characteristics of China's economic development. Energy is an indispensable important input factor for economic development. The resource protection policies proposed to reduce the amount of energy consumption would limit the speed of economic growth. At the same time, the fluctuations of energy prices and the shortage of energy will further restrict China's economic growth. Therefore, the Chinese government should correctly guide the transformation of the entire social production mode, improve the efficiency of energy use, reduce environmental pollution in the process of energy consumption, and reduce dependence on fossil fuel. Otherwise, the constraint effect of growing energy consumption on economic growth will continue to expand. In order to ensure sustainable high speed economic growth, it is necessary to gradually improve the current energy consumption structure, diversify channels of importing energy, research and develop new type of safe and clean energy to ensure adequate energy supply. To avoid the adverse impact of energy on economic growth, the Chinese government needs to adopt more comprehensive policies to transform the current production mode from extensive to intensive.

## References

- [1] Kraft J, Kraft A. On the relationship between energy and GNP [J]. *The Journal of Energy and Development*. 1978 Spring; 3(2): 401-403.
- [2] Akarca A T, Long II T V. Energy and employment: a time-series analysis of the causal relationship [J]. *Resources and Energy*. 1979 Oct–Nov; 2(2-3): 151-162.
- [3] Akarca A T, Long T V. On the relationship between energy and GNP: a reexamination [J]. *The Journal of Energy and Development*. 1980 Spring; 5(2): 326-331.
- [4] Eden S H, Hwang B K. The relationship between energy and GNP: further results [J]. *Energy Economics*. 1984 Jul; 6(3): 186-190.
- [5] Yu E S H, Choi J Y. The causal relationship between energy and GNP: an international comparison [J]. *The Journal of Energy and Development*. 1985 Spring; 10(2): 249-272.
- [6] Erol U, Eden S H. Time series analysis of the causal relationships between US energy and employment [J]. *Resources and Energy*. 1987 Jun; 9(1): 75-89.
- [7] Eden S H, Jin J C. Cointegration tests of energy consumption, income, and employment [J]. *Resources and Energy*. 1992 Sep; 14(3): 259-266.
- [8] Masih A M M, Masih R. Energy consumption, real income and temporal causality: results from a multi-country study based on cointegration and error-correction modelling techniques [J]. *Energy Economics*. 1996 Jul; 18(3): 165-183.
- [9] Masih A M M, Masih R. On the temporal causal relationship between energy consumption, real income, and prices: some new evidence from Asian-energy dependent NICs based on a multivariate cointegration/vector error-correction approach [J]. *Journal of Policy Modeling*. 1997 Aug; 19(4): 417-440.
- [10] Glasure Y U, Lee A R. Cointegration, error-correction, and the relationship between GDP and energy: The case of South Korea and Singapore [J]. *Resource and Energy Economics*. 1998 Mar; 20(1): 17-25.
- [11] Asafu-Adjaye J. The relationship between energy consumption, energy prices and economic growth: time series evidence from Asian developing countries [J]. *Energy Economics*. 2000 Dec; 22(6): 615-625.
- [12] Hondroyannis G, Lelos S, Papapetrou E. Energy consumption and economic growth: assessing the evidence from Greece [J]. *Energy Economics*. 2002 Jul; 24(4): 319-336.
- [13] Soytas U, Sari R. Energy consumption and GDP: causality relationship in G-7 countries and emerging markets [J]. *Energy Economics*. 2003 Jan; 25(1): 33-37.
- [14] Altinay G, Karagol E. Structural break, unit root, and the causality between energy consumption and GDP in Turkey [J]. *Energy Economics*. 2004 Nov; 26(6): 985-994.
- [15] Ang J B. CO<sub>2</sub> emissions, energy consumption, and output in France [J]. *Energy Policy*. 2007 Oct; 35(10): 4772-4778.
- [16] Ang J B. CO<sub>2</sub> emissions, research and technology transfer in China [J]. *Ecological Economics*. 2009 Aug; 68(10): 2658-2665.

- [17] Halicioglu F. An econometric study of CO<sub>2</sub> emissions, energy consumption, income and foreign trade in Turkey [J]. *Energy Policy*. 2009 Mar; 37(3): 1156-1164.
- [18] Jalil A, Mahmud S F. Environment Kuznets curve for CO<sub>2</sub> emissions: a cointegration analysis for China[J]. *Energy Policy*. 2009 Dec; 37(12): 5167-5172.
- [19] Apergis N, Payne J E. CO<sub>2</sub> emissions, energy usage, and output in Central America[J]. *Energy Policy*. 2009 Aug; 37(8): 3282-3286.
- [20] Apergis N, Payne J E. The emissions, energy consumption, and growth nexus: evidence from the commonwealth of independent states[J]. *Energy Policy*. 2010 Jan; 38(1): 650-655.
- [21] Hossain M S. Panel estimation for CO<sub>2</sub> emissions, energy consumption, economic growth, trade openness and urbanization of newly industrialized countries[J]. *Energy Policy*. 2011 Nov; 39(11): 6991-6999.

# Calculation and Analysis of Carbon Emissions for Green Buildings During Operation Period in Beijing

Guojian WANG<sup>1</sup>

*Beijing Institute of Residential Building Design & Research Co.,Ltd, Beijing, 100005, China*

**Abstract.** This study analyzed the carbon emissions for two-star green residential buildings in Beijing. During operation period, the carbon emissions were calculated for air-conditioning, heating, lighting, elevator, and domestic hot water systems. Furthermore, the carbon emissions for full life cycle are also given in this paper based on the statistical method of big data. The results show that the carbon emissions during operation and full life cycle periods were 17~21 and 21~27 kgCO<sub>2</sub>e/(m<sup>2</sup>·a), respectively. During operation period, the air-conditioning and heating systems accounted for 60%~70% of the total carbon emissions, 26%~30% and 4%~9% for lighting and elevator systems, and the carbon emissions were lowest for domestic hot water systems.

**Keywords:** Green building, carbon emission, operation period, calculation, analysis

## 1. Introduction

In response to global climate change, the Chinese government has set a target of reaching a peak in carbon dioxide emissions around 2030 and striving to be carbon neutrality by 2060. The extensively implement of green buildings is an effect way to reach the target of the peak carbon dioxide emissions and carbon neutrality.

Most of existed studies conducted by Chinese researchers on building carbon emissions during full or partial operation periods were based on life cycle assessment theory. Jin [1] analyzed the main influence factors on the variations of energy consumption in residential environments by using analytic hierarchy process (AHP), based on the established model related to energy-saving building, the concept of building carbon emission per person was also put forward. Liu [2] combined computational simulation and mathematical model to quantitatively analyze the emission-reduction effect in low-carbon residential building in hot summer and warm winter area respectively during construction and operation periods. Verbeeck et al. [3] calculated and analyzed the energy consumption and carbon emissions of full life cycle for 5 residential buildings with different building structures in Belgium, then the carbon emission list and model were further established. Gustavsson et al. [4] utilized emission coefficients to calculate the carbon emissions in every stage of buildings' full life cycle, and took a wooden residential building as an example to determine the energy consumption and carbon

---

<sup>1</sup>Corresponding author: Guojian Wang, Beijing Institute of Residential Building Design & Research Co., Ltd, Beijing, 100005, China; E-mail: [wgj86365413@126.com](mailto:wgj86365413@126.com).

emissions.

Unfortunately, the factors that influence building energy consumption such as heating load, cooling load, and power consumption were more emphasized in the processes of traditional design and consultant, while the carbon emissions factors were not further analyzed, leading to relatively less known on carbon emissions for green buildings. Therefore, this paper calculated and analyzed the monitored data of carbon emissions for air conditioning, heating, lighting, elevator, and domestic hot water systems during operation period in 3 green buildings. The relationship of carbon emissions between operation period and full life cycle period was obtained by using the statistical method of big data, furthermore, the carbon intensity for full life cycle can be acquired. This study aims to provide basic data and references for relevant studies.

## 2. Project Overview

3 two-star green residential buildings were selected to analyze, and the detailed information of the 3 buildings are summarized in table 1.

**Table 1.** Detailed information of 3 residential buildings.

No.	Location	Building area (m <sup>2</sup> )	Building height (m)	Floor	Air conditioning and heating systems	Performance of main equipment
1#	Fengtai district, Beijing	20811	79.8	28 floors above the ground and 4 floors underground	Cooling with wall-mounted split air conditioners, radiant floor heating system with central gas boiler	The energy efficiency ratio of air conditioner is 3.4, the thermal efficiency of the boiler is 90%
2#	Shunyi district, Beijing	4802	27.6	9 floors above the ground and 2 floors underground	Cooling with wall-mounted split air conditioners, radiant floor heating system with wall-mounted gas boiler	The energy efficiency ratio of air conditioner is 3.4, the thermal efficiency of the boiler is 84%
3#	Pinggu district, Beijing	9668	48	16 floors above the ground and 2 floors underground	Cooling with wall-mounted split air conditioners, radiant floor heating system with municipal heating	The energy efficiency ratio of air conditioner is 3.4, the supply (return) water for municipal heating and residents are 110 (70) °C and 50 (40) °C

### 3. Calculation and Analysis

#### 3.1. Determination of Boundary Condition

The cooling and heating loads of the 3 buildings were simulated by DeST software. The heating period was from 15th November to 15th March of the next year, and the cooling period was from first June to 31th August. The indoor design parameters are shown in table 2.

**Table 2.** Indoor design parameters.

Room type	Summer		Winter	
	Temperature (°C)	Relative humidity (%)	Temperature (°C)	Relative humidity (%)
Living room	26	65	20	30
Bedroom	26	65	20	30
Kitchen and bathroom	26	65	20	-
Stairwell	-	-	-	-

#### 3.2. Energy Consumption Calculations for Different Buildings

##### (1) Building 1#

###### a) Calculations of energy consumption for air conditioning and heating systems

According to the simulated results of building 1#, the cooling and heating loads were 293.55 MWh and 497.92 MWh, respectively, and the convey efficiency was 0.93 for outdoor pipes [5]. The total energy consumption for cooling with air conditioners and heating was 86.33 MWh/a and 595 MWh/a, respectively, corresponding to 4.71 kWh/(m<sup>2</sup>·a) power and 3.4 Nm<sup>3</sup>/(m<sup>2</sup>·a) natural gas. The carbon emission factor was 2.16 kgCO<sub>2</sub>e/Nm<sup>3</sup> for natural gas [6], and the carbon emission factor was 0.604 kgCO<sub>2</sub>e/kWh for electric power in Beijing [7]. Therefore, the total carbon emission for cooling and heating systems in building 1# was 10.18 kgCO<sub>22·a).</sub>

###### b) Calculation of energy consumption for lighting system

Based on the *Standard for building carbon emission calculation GBT51366-2019*, the current limitation values for lighting power and lighting time in the 3 residential building were determined and shown in table 3.

**Table 3.** The limitations for lighting power and lighting time in the 3 residential buildings.

Room type	Lighting time within a month (h)	Lighting power density (W/m <sup>2</sup> )
Living room	165	6
Bedroom	135	6
Kitchen	96	6
Bathroom	165	6
Public space such as corridor, etc.	5	15

The calculated results showed that the total energy consumption for lighting system in building 1# was 8.66 kWh/(m<sup>2</sup>·a) throughout a year, and the corresponding carbon emission was 5.23 kgCO<sub>2</sub>e /(m<sup>2</sup>·a).

###### c) Calculation of energy consumption for elevator

According to the predictive model of energy consumption for elevator which has been provided in the *Specification for electric lifts GB/T10058-2009*, the energy

consumption for elevator can be calculated by equation (1).

$$E_{el} = (K_1 \times K_2 \times K_3 \times H \times F \times P) / (V \times 3600) + Est \quad (1)$$

Where  $E_{el}$  is the energy consumption for elevator throughout a year, kWh/a;  $K_1$  is the driving coefficient,  $K_1=1.6$  (driven by ACVV system),  $K_1=1.0$  (driven by VVF system),  $K_1=0.6$  (driven by VVF system with energy feedback);  $K_2$  is the average travel distance coefficient,  $K_2=1.0$  (two floors),  $K_2=0.5$  (more than two floors with single elevator or two elevators parallel operation),  $K_2=0.3$  (no less than two elevators and with group control);  $K_3$  is the average load coefficient of elevator car,  $K_3=0.35$ ;  $H$  is the longest travel distance, m;  $F$  is the yearly number of starting times, based on the references on normal residential building, herein,  $F$  is about 146000 times;  $P$  is rated power, kW;  $V$  is rated speed, m/s;  $Est$  is the yearly total energy consumption of idle state, which is about account for 70% of total energy consumption.

Building 1# had 2 dwelling units and 4 elevators, all elevators were driven by VVF system. For each elevator, the rated load was 825kg, the longest travel distance was 80m, and the rated speed was 2m/s. According to equation (1), the yearly energy consumption for the elevators in building 1# can be calculated, corresponding to 2.53 kWh/(m<sup>2</sup>·a) and 1.53 kgCO<sub>2</sub>e /(m<sup>2</sup>·a).

#### d) Calculation of energy consumption for domestic hot water system

The energy consumption of domestic hot water system in residential building should be analyzed based on the actual operation condition and that can be calculated by equations (2) and (3):

$$Q_{rp} = 4.178 \frac{mq_r (t_r - t_l) \rho_r}{1000} \quad (2)$$

$$Q_r = T Q_{rp} \quad (3)$$

Where  $Q_r$  is the yearly energy consumption of domestic hot water system, (kWh/a);  $Q_{rp}$  is hourly energy consumption of domestic hot water system, (kW/h);  $m$  is the unit of water consumption;  $q_r$  is the limitation of domestic hot water for each people (L/person), which can be determined by the *Standard for water saving design in civil building GB50555-2010*;  $\rho_r$  is the density of domestic hot water (kg/L);  $t_r$  is the design temperature of domestic hot water (°C);  $t_l$  is the design temperature of cold water (°C).

20 L/person domestic hot water was used in this study. According to the *Standard for design of building water supply and drainage GB50015-2019*, the design temperature of cold water is 4 °C. Therefore, the yearly energy consumption of domestic hot water system was 399.4 kWh/a, corresponding to 0.022 kWh/(m<sup>2</sup>·a), and the carbon emission of domestic hot water in building 1# was 0.013 kgCO<sub>2</sub>e /(m<sup>2</sup>·a).

In summary, the carbon emission for building 1# during operation period was 16.95 kgCO<sub>2</sub>e /m<sup>2</sup>.

## (2) Building 2#

### a) Calculations of energy consumption for air conditioning and heating systems

According to the simulated results of building 2#, the cooling and heating loads for building 2# were 61.52 MWh and 149.34 MWh, respectively, and the convey efficiency was 0.93 for outdoor pipes. The total energy consumption for cooling and heating systems were 18.09 MWh/a and 177.78 MWh/a, respectively, corresponding to 4.18

$\text{kWh}/(\text{m}^2 \cdot \text{a})$  electric power and  $4.29 \text{ Nm}^3/(\text{m}^2 \cdot \text{a})$  natural gas. The carbon emission factor for natural gas was  $2.16 \text{ kgCO}_2\text{e}/\text{Nm}^3$ , and the carbon emission factor for electric power in Beijing was  $0.604 \text{ kgCO}_2\text{e}/\text{kWh}$ . Therefore, the total carbon emission of cooling and heating systems for building 2# was  $11.79 \text{ kgCO}_2\text{e}/(\text{m}^2 \cdot \text{a})$ .

#### b) Calculation of energy consumption for lighting system

The calculated results showed that the total energy consumption for lighting system in building 2# was  $9.83 \text{ kWh}/(\text{m}^2 \cdot \text{a})$  throughout a year, and the corresponding carbon emission was  $5.94 \text{ kgCO}_2\text{e}/(\text{m}^2 \cdot \text{a})$ .

#### c) Calculation of energy consumption for elevator

Building 2# had 2 dwelling units and 2 elevators, both of the elevators were driven by VVF system. For each elevator, the rated load was 1000 kg, the longest travel distance was 30m, and the rated speed was 2m/s. According to equation (1), the yearly energy consumption for the elevators in building 2# can be calculated, corresponding to  $2.01 \text{ kWh}/(\text{m}^2 \cdot \text{a})$  and  $1.21 \text{ kgCO}_2\text{e}/(\text{m}^2 \cdot \text{a})$ .

#### d) Calculation of energy consumption for domestic hot water system

The yearly energy consumption of domestic hot water system in building 2# was  $59.9 \text{ kWh/a}$ , corresponding to  $0.014 \text{ kWh}/(\text{m}^2 \cdot \text{a})$ , and the carbon emission of domestic hot water in building 2# was  $0.0085 \text{ kgCO}_2\text{e}/(\text{m}^2 \cdot \text{a})$ .

In summary, the total carbon emission for building 2# during operation period was  $18.95 \text{ kgCO}_2\text{e}/\text{m}^2$ .

### (3) Building 3#

#### a) Calculation of energy consumption for air conditioning and heating systems

According to the simulated results of building 3#, the cooling and heating loads were 155.58 MWh and 262.82MWh, respectively, and the convey efficiency was 0.93 for outdoor pipes. The energy consumption for cooling with air conditioners and heating were 45.8 MWh/a and 105.0 MWh/a, respectively, corresponding to  $4.73 \text{ kWh}/(\text{m}^2 \cdot \text{a})$  electric power. The carbon emission factor of thermal energy was  $110 \text{ kgCO}_2\text{e}/\text{GJ}$  [8]. The total carbon emission of cooling and heating systems in building 3# was  $14.4 \text{ kgCO}_2\text{e}/(\text{m}^2 \cdot \text{a})$ .

#### b) Calculation of energy consumption for lighting system

The calculated results showed that the total energy consumption for lighting system throughout a year in building 3# was  $9.01 \text{ kWh}/(\text{m}^2 \cdot \text{a})$ , and the corresponding carbon emission was  $5.42 \text{ kgCO}_2\text{e}/(\text{m}^2 \cdot \text{a})$ .

#### c) Calculation of energy consumption for elevator

Building 3# had 2 dwelling units and 4 elevators, all elevators were driven by VVF system. For each elevator, the rated load was 1000 kg, the longest travel distance was 48m, and the rated speed was 2m/s. According to equation (1), the yearly energy consumption for the elevators in building 1# can be calculated, corresponding to  $1.518 \text{ kWh}/(\text{m}^2 \cdot \text{a})$  and  $0.91 \text{ kgCO}_2\text{e}/(\text{m}^2 \cdot \text{a})$ .

#### d) Calculation of energy consumption for elevator domestic hot water system

The yearly energy consumption of domestic hot water system in building 3# was  $228.22 \text{ kWh/a}$ , corresponding to  $0.023 \text{ kWh}/(\text{m}^2 \cdot \text{a})$ , and the carbon emissions of domestic hot water in building 3# was  $0.014 \text{ kgCO}_2\text{e}/(\text{m}^2 \cdot \text{a})$ .

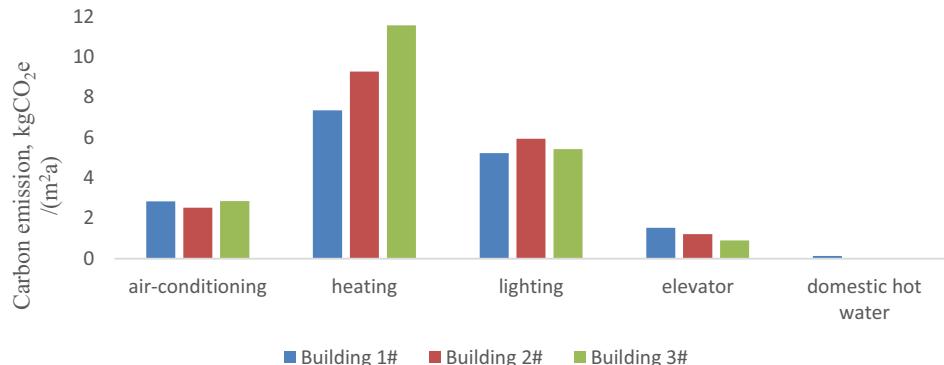
In summary, the total carbon emission in building 1# during operation period was  $20.74 \text{ kgCO}_2\text{e}/\text{m}^2$ .

### 3.3. Comparison and Analysis of the 3 Buildings

#### a) Summary of the carbon emissions in 3 buildings

The calculated results of different buildings are summarized in figure 1 and table 4~table 6, the results showed that:

- 1) The carbon emissions of different buildings ranged from 17~21 kgCO<sub>2</sub>e/(m<sup>2</sup>·a).
- 2) The cooling and heating systems accounted for 60%~70% of total carbon emissions, and it was 26%~30% for lighting system, 4%~9% for elevators, while the proportion of domestic hot water system was the lowest.



**Figure 1.** The carbon emissions for different systems in 3 buildings.

**Table 4.** The calculated results of building 1#.

No.	System	kgCO <sub>2</sub> e/(m <sup>2</sup> ·a)	Proportion
1	Air conditioning	2.84	16.76%
2	Heating	7.34	43.30%
3	Lighting	5.23	30.86%
4	Elevator	1.53	9.03%
5	Domestic hot water	0.13	0.77%
Total		16.95	

**Table 5.** The calculated results of building 2#.

No.	System	kgCO <sub>2</sub> e/(m <sup>2</sup> ·a)	Proportion
1	Air conditioning	2.52	13.30%
2	Heating	9.27	48.92%
3	Lighting	5.94	31.35%
4	Elevator	1.21	6.39%
5	Domestic hot water	0.0085	0.044%
Total		18.95	

**Table 6.** The calculated results of building 3#.

No.	System	kgCO <sub>2</sub> e/(m <sup>2</sup> ·a)	Proportion
1	Air conditioning	2.85	13.74%
2	Heating	11.55	55.69%
3	Lighting	5.42	26.13%
4	Elevator	0.91	4.38%
5	Domestic hot water	0.014	0.068%
Total		20.74	

### b) Analysis of carbon emissions of air conditioning and heating systems

The carbon emissions of different heating sources are shown in table 7. The carbon emissions present regional central gas boiler<wall-mounted gas boiler< municipal heating. The thermal sources of municipal heating are mainly provided by combined heat and power (CHP) or central boiler, to reduce the carbon emissions, the use of clean energy and the types of energy should be enhanced. The carbon emissions of air conditioning system mainly ranged from 2 to 3 kgCO<sub>2</sub>e/(m<sup>2</sup>·a).

**Table 7.** Carbon emissions of air conditioning and heating systems.

No.	Air conditioning and heating systems	Performance of main equipment	Carbon emission kgCO <sub>2</sub> e/(m <sup>2</sup> ·a)
1	Radiant floor heating system with central gas boiler	The thermal efficiency of the boiler is 90%	7.34
2	Cooling with wall-mounted split air conditioners, radiant floor heating system with wall-mounted gas boiler	the thermal efficiency of the boiler is 84%	9.27
3	Cooling with wall-mounted split air conditioners, radiant floor heating system with municipal heating	The supply (return) water for municipal heating and residents were 110 (70) °C and 50 (40) °C	11.55

### c) Carbon emissions of full life cycle

Referring to the building carbon emissions reported by different studies [9-13], the carbon emissions during operating period accounted for 70%~90% of total carbon emissions within full life cycle. In terms of this study, the building carbon emissions of full life cycle were 21~27 kgCO<sub>2</sub>e/(m<sup>2</sup>·a).

## 4. Conclusions

- (1) The buildings' carbon emissions during operation period and full life cycle were 17~21 and 21~27 kgCO<sub>2</sub>e/(m<sup>2</sup>·a), respectively.
- (2) The cooling and heating systems accounted for 60%~70% of total carbon emissions, and it was 26%~30% for lighting system, 4%~9% for elevators, while the proportion of domestic hot water system was the lowest, less than 1%.
- (3) The carbon emissions of air conditioning system during operation period mainly ranged from 2 to 3 kgCO<sub>2</sub>e/(m<sup>2</sup>·a).
- (4) The carbon emissions present regional central gas boiler<wall-mounted gas boiler< municipal heating.

## List of notations

- $E_{el}$  is the energy consumption for elevator throughout a year  
 $K_1$  is the driving coefficient of elevator  
 $K_2$  is the average travel distance coefficient of elevator  
 $K_3$  is the average load coefficient of elevator car of elevator  
 $H$  is the longest travel distance of elevator  
 $F$  is the yearly number of starting times of elevator  
 $P$  is rated power of elevator  
 $V$  is rated speed of elevator  
 $E_{st}$  is the yearly total energy consumption of idle state  
 $Q_r$  is the yearly energy consumption of domestic hot water system  
 $Q_{rp}$  is hourly energy consumption of domestic hot water system  
 $m$  is the unit of water consumption  
 $q_r$  is the limitation of domestic hot water for each people  
 $\rho_r$  is the density of domestic hot water  
 $t_r$  is the design temperature of domestic hot water  
 $t_l$  is the design temperature of cold water

## References

- [1] Jin J. Energy efficiency in residential building research based on carbon emissions. 2014; Changan University.
- [2] Liu J. Quantified study on CO<sub>2</sub> reduction effect of low-carbon housing technology in hot summer and cold winter zone. 2014. Huazhong University of Science and Technology.
- [3] Verbeeck G, Hens H. Life cycle inventory of buildings: A calculation method. Building and Environment. 2010; 45(4):1037-41.
- [4] Gustavsson L, Joellsson A, Sathre R. Life cycle primary energy use and carbon emission of an eight-storey wood-framed apartment building. Energy and Buildings. 2010; 42(2):230-42.
- [5] Design standard for energy efficiency of residential buildings DB11/891-2020.
- [6] Technical Guidelines for Assessment of Low-carbon Building Operation DB11/T 1420—2017.
- [7] Requirements for carbon dioxide emission accounting and reporting Power generation enterprises DB11/T 1781—2020.
- [8] Requirements for carbon dioxide emission accounting and reporting Heat production and supply enterprises DB11/T 1784—2020.
- [9] Peng B 2012 Case Study on Life Cycle Energy Consumption and CO<sub>2</sub> Emissions of Green Buildings Tsinghua University.
- [10] Chen S, Cui D, Zhang H. Calculation of carbon dioxide emissions from buildings and case study. Journal of Beijing University of Technology. 2016; 42(04):594-600.
- [11] Wang Y, Yang X, Yan H, Zhang Y, Li J. Carbon emission accounting for buildings based on whole life cycle: A case study of reconstruction project at college in guangzhou. Journal of Engineering Management. 2017; 31(03):19-24.
- [12] Li J, Liu Y. The carbon emission accounting model based on building lifecycle. Journal of Engineering Management. 2015; 29(04):12-16.
- [13] Yu J, Zhang Y, Qi S. Study on carbon emission characteristics and emission reduction strategies of public buildings in the hot summer and warm winter areas throughout their life cycles—Take Xiamen city as an example. Building Science. 2014; 30(02): 13-8.

# Location Optimization of Hydrogen Refueling Stations in Hydrogen Expressway Based on Hydrogen Supply Chain Cost

Dongxing WANG<sup>a</sup>, Zhe WANG<sup>a,b,1</sup>, Fenghui HAN<sup>a,b</sup>, Fan ZHAO<sup>a</sup>, Yulong JI<sup>a</sup>

<sup>a</sup>*Marine Engineering College, Dalian Maritime University, Dalian, 116026, China*

<sup>b</sup>*National Center for International Research of Subsea Engineering Technology and Equipment, Dalian Maritime University, Dalian, 116026, China*

**Abstract.** Hydrogen energy is regarded as an important way to achieve carbon emission reduction. This paper focuses on the combination of the design of the hydrogen supply chain network and the location of hydrogen refueling stations on the expressway. Based on the cost analysis of the hydrogen supply chain, a multi-objective model is developed to determine the optimal scale and location of hydrogen refueling stations on the hydrogen expressway. The proposed model considers the hydrogen demand forecast, hydrogen source selection, hydrogen production and storage and transportation, hydrogen station refueling mode, etc. Taking Dalian City, China as an example, with offshore wind power as a reliable green hydrogen supply to select the location and capacity of hydrogen refueling stations for the hydrogen energy demonstration section of a certain expressway under multiple scenarios. The results of the case show that 4 and 5 stations are optimized on the expressway section respectively, and the unit hydrogen cost is \$14.3 /kg H<sub>2</sub> and \$11.8 /kg H<sub>2</sub> respectively, which are equal to the average hydrogen price in the international range. The optimization results verify the feasibility and effectiveness of the model.

**Keywords:** hydrogen refueling stations, hydrogen supply chain, hydrogen expressway, location model, cost analysis

## 1. Introduction

Hydrogen has become an important direction of future energy development. At present, hydrogen fuel cell vehicle (HFCV) is the most efficient and production-ready hydrogen technology available [1], and becomes a main direction of hydrogen energy development. However, the lack of complete hydrogen road network and sufficient supporting hydrogen refueling stations (HRSs) are the obvious problems restricting the commercial development of HFCVs. It is estimated that by 2030, there will be 15 million HFCVs and 15,000 HRSs in the world [2]. Despite the rapid construction of global HRSs, the infrastructure of HRSs still faces many challenges [3].

Reasonable location and layout of HRSs can not only make up for the shortage of hydrogen refueling infrastructure, but also have certain practical significance for the construction of hydrogen energy road network. Scholars in various countries have carried

---

<sup>1</sup> Corresponding Author: Zhe Wang, Marine Engineering College, Dalian Maritime University, Dalian, 116026, China; wang.zhe@dlmu.edu.cn, +86-0411-84729038.

out relevant studies on the location of hydrogen refueling infrastructure such as HRSs. Some people choose the location of HRSs based on the point demand. Frade et al. [4] applied the set coverage model of point demand to study the location of Avenidas HRSs, which was implemented in Lisbon. Hadi [5] proposed a two-step modeling and solution to the large-scale maximum coverage location problem, and obtained the optimal resource allocation in downtown Pittsburgh. Nicholas et al. [6] combined the P-median model with geographic information system and applied it to the location of HRSs in California. Different from the point demand model applied by the above scholars, Hodgson et al. [7] applied the flow capture location model (FCLM) and combined the greedy algorithm to locate the site of the HRSs in Edmonton. Kuby et al. [8] applied the actual road network data of Florida to optimize the location problem of HRSs. In order to explore the interaction between the sales volume of HFCVs and the number of HRSs, Yushan et al. [9] solved the integrated optimization model using the capacity constrained FCLM algorithm and the solution process to optimize the location of HRSs. Most scholars focused on a single city or region. Onur et al. [10] studied the set covering constraints to ensure that each area will be covered by the minimum amount of hydrogen infrastructure. Lei et al. [11] combined the classic hydrogen supply chain network design model with the HRSs location planning model. Sungmi [12] proposed an optimization model to effectively determine when and where to build hydrogen production bases and HRSs in urban areas. However, there are few studies on the location of HRSs based on expressways. Besides, the construction of HRSs on expressways can refer to the research of charging stations to some extent [13, 14].

Based on the above the hydrogen analysis and the research of HRSs location, this paper focuses on the location of HRSs on the expressway, combines location problem with hydrogen supply chain network design, aims to minimize the total hydrogen supply chain network cost, and develops a multi-objective model based on linear model to determine the optimal size and location of HRSs.

## 2. Construction of mathematical model for location of hydrogen refueling stations

### 2.1. Modeling of Location Factors for Hydrogen Refueling Stations

The hydrogen supply chain mainly includes hydrogen production, storage, transportation, and hydrogen refueling station. Therefore, the total cost of hydrogen supply chain  $C$  is composed of hydrogen production cost  $C_p$ , hydrogen storage and transportation cost  $C_{SD}$ , and hydrogen refueling stations cost  $C_H$ , which is expressed as follows:

$$C = C_p + C_{SD} + C_H \quad (1)$$

The hydrogen production cost is related to the production method of hydrogen and the capacity of HRS, which is shown in the equation (2):

$$C_p = \sum_{i \in N} \sum_{j \in M} p_j \times w_{ij} \quad (2)$$

where  $C_p$  is hydrogen production cost,  $p_j$  is the unit hydrogen production cost of hydrogen source  $j$  (\$/kg H<sub>2</sub>), and  $w_{ij}$  is the mass of hydrogen supplied by hydrogen source  $j$  for HRS  $i$  (kg). The hydrogen storage and transportation cost are related to the

storage and transportation mode of hydrogen, the transportation distance of hydrogen and the capacity of HRS, and its expression is shown in the equation (3):

$$C_{SD} = \sum_{i \in N} \sum_{j \in M} c_{ij} \times w_{ij} \quad (3)$$

where  $C_{SD}$  is hydrogen storage and transportation cost,  $c_{ij}$  is the unit hydrogen storage and transportation cost (\$/kg H<sub>2</sub>) of hydrogen source  $j$  supplying HRS  $i$ , which is calculated according to the distance  $d_{ij}$  (km) between hydrogen source  $j$  and HRS  $i$ , cost index of transportation vehicle  $A$  [15] and hydrogen storage and transportation cost index  $B$ . The expressions are shown in the equation (4):

$$c_{ij} = A \times d_{ij} + B \quad (4)$$

The hydrogen refueling station cost is related to the capacity of HRS as follows:

$$C_H = \sum_{i \in N} \sum_{j \in M} c_{hi} \times w_{ij} \quad (5)$$

where  $C_H$  is hydrogen refueling stations cost,  $c_{hi}$  is the unit cost of HRS (\$/kg H<sub>2</sub>).

## 2.2. Comprehensive Mathematical Model and Hypothesis

Combined with the above factors of the HRSs location, the model is constructed with the goal of minimizing the total cost of hydrogen supply chain, as shown in the equation (6):

$$\text{Minimize } C = \sum_{i \in N} \sum_{j \in M} p_j \times w_{ij} + \sum_{i \in N} \sum_{j \in M} (A \times d_{ij} + B) \times w_{ij} + \sum_{i \in N} \sum_{j \in M} c_{hi} \times w_{ij} \quad (6)$$

with the constraint conditions  $F$  below:

$$F = \begin{cases} \sum_{i \in N} w_{ij} \leq L_j, \sum_{j \in M} w_{ij} \leq w_i, \sum_{i \in N} w_i \geq D \\ \sum_{j \in M} L_j \geq \sum_{i \in N} w_i, d_{i,i+1} \leq 200 \end{cases} \quad (7)$$

where  $L_j$  is the maximum capacity of hydrogen source  $j$  (kg H<sub>2</sub>/d),  $w_i$  is the capacity of HRS  $i$  (kg H<sub>2</sub>/d),  $D$  is the total hydrogen demand of expressway section (kg H<sub>2</sub>/d), and  $d_{i,i+1}$  is the distance between adjacent HRSs. Constraint (7) are used to constrain hydrogen demand and station distance of the model, respectively. Based on the above model construction, the hypothesis and applicable conditions are described as follows: 1) expressway is directional and one-way irreversible; 2) when the driver misses a refueling station, he needs to drive to the next station to complete the refueling behavior; 3) when HFCVs are hydrogenated, the hydrogen capacity shall not be less than 50%.

## 3. Location Selection and Optimization Scheme of Hydrogen Refueling Stations

### 3.1. Key Parameters and Applicable Conditions

Based on the above mathematical model, the hydrogen supply chain network of the 300 km section of the expressway from Dalian to Dandong city was designed, and the

location of HRSs was optimized. As a key input, changes in hydrogen demand have a significant impact on the structure of the hydrogen supply chain [16]. Estimations of hydrogen demand are mainly based on the key parameters such as average daily vehicle mileage per kilometer of expressways, fuel economy of HFCVs and market share of HFCVs. The main parameters and total hydrogen demand are summarized in Table 1.

**Table 1.** Summary of key parameters of hydrogen demand in expressway section.

Parameters	Value
Total expressway length (km)	300
The market share of HFCVs (%)	5
Roewe950 single hydrogenation (kg)	4.2
Roewe950 Range (km)	430
Average daily vehicle mileage per km of expressway (veh.km/km/d)	20,000
The total hydrogen demand (kg H <sub>2</sub> /d)	2,930

Based on the characteristics of the fluctuation of offshore wind energy, the method of water electrolysis for hydrogen production in this case is PEM [17]. To estimate the unit cost of hydrogen production, the case study used Hydrogen Analysis Production Models (H2A) developed by the National Renewable Energy Laboratory (NREL) [18]. The relevant information of PEM electrolytic cell and the result of the total hydrogen production cost are summarized in Table 2.

**Table 2.** Summary of economic parameters of PEM and cost result based on H2A.

Parameters	Value
Hydrogen production scale of hydrogen source (kg H <sub>2</sub> /d)	5,000
Cost of equipment (\$/kW)	233.61
Power consumption (kWh/kg H <sub>2</sub> )	51.3
Electricity Price (Offshore Wind Power) (\$/kWh)	0.04
Stack life (years)	10
Total cost of hydrogen production (\$/kg H <sub>2</sub> )	3.02

The cost of hydrogen storage, transportation and HRSs are evaluated based on the Hydrogen Delivery Scenario Analysis Model (HDSAM) developed by Argonne National Laboratory [19]. High-pressure gas and hydrogen transportation of long tube trailers, cascade storage and refueling pressure of 35 MPa at HRSs are considered as key inputs and assumptions in the scenario of HDSAM, and the results are shown in Table 3.

**Table 3.** The summary of results based on HDSAM calculation.

Daily refueling capacity (kg H <sub>2</sub> /d)	500	800	1,000
Compressed H2 terminal cost (\$/kg H <sub>2</sub> )	2.2	2.2	2.2
Cascade storage cost (\$/kg H <sub>2</sub> )	0.49	0.49	0.49
Compressed H2 truck cost (\$/kg H <sub>2</sub> )	1.82	1.43	1.30
Hydrogen storage and transportation cost index B	4.5	4.11	3.98
Hydrogen refueling station cost (\$/kg H <sub>2</sub> )	2.92	3.04	2.69

### 3.2. Optimization Scenario Definition

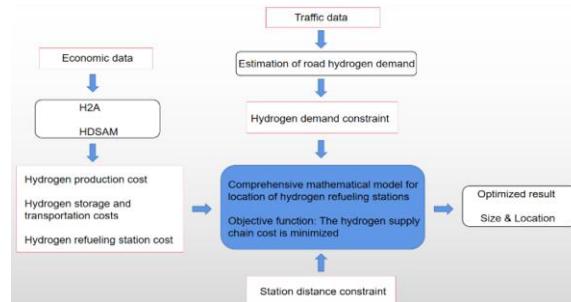


**Figure 1.** Schematic figure of location about hydrogen sources and transportation nodes.

On the premise of satisfying the hydrogen demand in the expressway section, two scenarios of single hydrogen source (Huayuankou) and single capacity HRS (800 kg H<sub>2</sub>/d) and double hydrogen source (Huanyuankou and Donggang) and multi-capacity HRS (500 kg H<sub>2</sub>/d, 800 kg H<sub>2</sub>/d, 1,000 kg H<sub>2</sub>/d) are selected for optimization. The scale of hydrogen production from hydrogen source is 5,000 kg H<sub>2</sub>/d. The expressway, location distribution of hydrogen sources and transportation nodes is shown in Figure 1.

### 3.3. Optimization Solution

According to the comprehensive mathematical model of hydrogen refueling location and the definition of optimization scenario, the data in Table 1-3 are taken as the key inputs of the model. Combined with the calculation framework Figure 2, the program is written with the help of MATLAB mathematical tools, and the PSO is used to solve the model. Specifically, the case calculates the total hydrogen demand as the demand constraint to determine the number and capacity of HRSs. And the location of HRSs is determined by the driving distance limit. With the help of H2A and HDSAM, the unit cost of hydrogen production, storage, transportation and refueling process is estimated. The above results are used as the inputs of the model, and then the model is optimized and solved.



**Figure 2.** The calculation framework of model.

## 4. Optimization Results and Discussion

The algorithm convergence results under the two optimization scenarios are shown in Figure 3-4. The program running results are sorted out to obtain the optimization results of the location and capacity of the HRSs under the two scenarios, as shown in Table 4. By combining the scenario optimization results with the distribution of hydrogen sources and transportation nodes, the results can be obtained as shown in Figure 5.

According to the results of scenario optimization in Table 4, a total of 4 stations are optimized in S1 on the Dan-Da section of He-Da expressway with a total length of about 300 km. The total daily cost of hydrogen supply chain is \$45,900, the locations of HRSs are set at 44.8 km, 105.2 km, 165.3 km and 225.9 km, respectively, and the unit hydrogen cost is about \$14.3 /kg H<sub>2</sub>. In S2, a total of 5 stations are optimized and built in the whole section, which are built at 54 km, 105 km, 151 km, 222 km and 262 km respectively, and the corresponding daily refueling capacities are 500 kg H<sub>2</sub>/d, 800 kg H<sub>2</sub>/d, 500 kg H<sub>2</sub>/d, 1,000 kg H<sub>2</sub>/d and 800 kg H<sub>2</sub>/d respectively. The total cost of hydrogen supply chain is \$42,500, and the unit hydrogen cost is about \$11.8 /kg H<sub>2</sub>. In both scenarios, unit hydrogen costs are in line with the international average selling price of hydrogen.

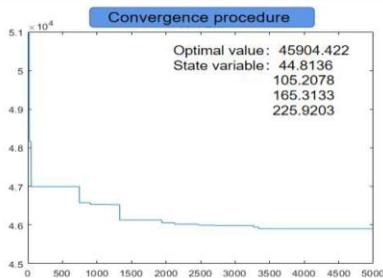


Figure 3. Scenario 1 convergence rendering.

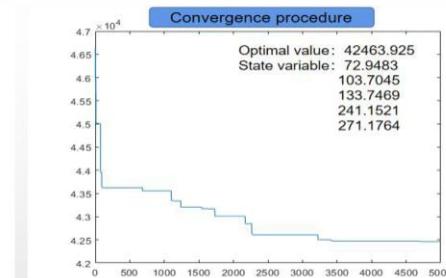


Figure 4. Scenario 2 convergence rendering.

Table 4. Summary of hydrogen sources and transportation nodes.

Scenario	Daily refueling capacity (kg H <sub>2</sub> /d)	Location of HRS (km)	Hydrogen source	Total cost (\$/d)	Unit cost (\$/kg H <sub>2</sub> )		
S 1	800	44.8	Huayunakou	45,904.422	14.3		
		105.2					
		165.3					
		225.9					
S 2	500	72.9	Huayuankou	42,463.925	11.8		
	800	103.7					
	500	133.7					
	1,000	241.1	Donggang				
	800	271.1					



Figure 5. Scenario 1-2: Location optimization result of HRSs.

Different from the difference in economy in Table 4, Figure 5 more intuitively show the distribution of HRSs on the whole expressway section. As can be seen from Figure 5, compared with the situation that the 5 HRSs are optimized in S2, the positions of the 4 HRSs in S1 are more evenly distributed in the whole expressway section, while the distance between the station 3 and 4 in S2 is larger. Moreover, in S2, HRSs are distributed more obviously centered on transportation nodes. However, in the two scenarios, the location and capacity of HRSs are in line with the relevant constraint settings of the model. Therefore, the optimization results in the two scenarios meet the requirements and the scheme is relatively ideal.

## 5. Conclusion

This paper summarizes the status of the hydrogen industry, starts from the hydrogen expressway and HRSs layout, combines hydrogen supply chain network design and HRSs location, analyses the global to the entire network with hydrogen, and proposes a minimum cost for the purpose of hydrogen expressway with HRSs location universal

model, in order to promote the development of the hydrogen industry. The proposed method, based on the hydrogen supply chain network with the cost of HRSs location model, can be applied to help an area with hydrogen of an expressway network transform traditional car early transition plan for HFCVs, but with the large-scale development of the hydrogen industry, the model should be combined with the application scenario to make improvement and adjustment, in order to apply to a wider variety of scenarios with more comprehensive and accurate optimization results.

## Funding

This work was funded by the National Natural Science Foundation of China (51906026), China Postdoctoral Science Foundation (2020M680928), Natural Science Foundation of Liaoning Province (2020-BS-067) and 111 Project (B18009).

## References

- [1] Jennifer K, Sam S, Thomas H.B. Review of transportation hydrogen infrastructure performance and reliability. *Int J Hydrogen Energy*. 2019; 44: 12010-12023.
- [2] Apostolou, D and Xydis G. A literature review on hydrogen refuelling stations and infrastructure. Current status and future prospects. *Sustain. Energy Rev.* 2019 Jul; 113: 109292.
- [3] Hardman S, Shiu E, Steinberger-Wilckens R, Turrentine T. Barriers to the adoption of fuel cell vehicles: a qualitative investigation into early adopters attitudes. *Transport. Res. Pol. Pract.* 2017; 95: 166-82.
- [4] Frade I, Ribeiro A, Goncalves G and Antunes AP. Optimal location of charging stations for electric vehicles in a neighborhood in Lisbon, Portugal. *Transp Res Rec*. 2011; 2252(1): 91-8.
- [5] Hajari H and Karimi H A. A Method for Large-Scale Resource Allocation in Smart Cities. 2018.
- [6] Nicholas MA, Handy SL and Sperling D. Using geographic information systems to evaluate siting and networks of hydrogen stations. *Transp Res Rec*. 2004; 1880(1): 126-34.
- [7] Hodgson MJ, and Rosing KE. Applying the flow-capturing location-allocation model to an authentic network: Edmonton, Canada. *European Journal of Operational Research*. 2007; 90(3): 427-43.
- [8] Kuby M, Lines L, Schultz R, Xie Z, Kim JG and Lim S. Optimization of hydrogen stations in Florida using the flow-refueling location model. *Int J Hydrogen Energy*. 2009; 34(15): 6045-64.
- [9] Yushan L, Fengming C and Lefei L. An integrated optimization model for the location of hydrogen refueling stations. *Int J Hydrogen Energy*. 2018; 43: 19636-49.
- [10] Onur D, Elifcan G, Ebru Y, Rizvan E. Amathematical programming model for facility location optimization of hydrogen production from enewable energy sources. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 2020.
- [11] Lei L, Herve M and Marie-Ange M. Integrated optimization model for hydrogen supply chain network design and hydrogen fueling station planning. *Computers and Chemical Engineering*. 2019; 134: 106683.
- [12] Sungmi B, Eunhan L and Jinil H. Multi-Period Planning of Hydrogen Supply Network for Refuelling Hydrogen Fuel Cell Vehicles in Urban Areas. *Sustainability*. 2020; 12(10): 4114.
- [13] Pevec D, Babic J and Kayser M A. A data-driven statistical approach for extending electric vehicle charging infrastructure. *International Journal of Energy Research*. 2018; 42(4): 3102-20.
- [14] Qunshan Z, Scott B.K, Fan X and Michael J.K. A multi-scale framework for fuel station location: From highways to street intersections. *Transportation Research Part D*. 2019; 74: 48-64.
- [15] Stevens M.B, Fowler M., Elkamel A, Elhedhli S. Macro-level optimized deployment of an electrolyser-based hydrogen refueling infrastructure with demand growth. *Eng. Optim.* 2018; 40(10): 955-67.
- [16] Lei L, Hervé M and Marie-Ange M. Hydrogen supply chain network design: An optimization-oriented review. *Renewable and Sustainable Energy Reviews*. 2019; 103: 342-60.
- [17] Zexue D and Xuhong M. Development of water electrolysis technology and its application in green hydrogen production. *Petroleum Processing and Petrochemicals*. 2021; 52(2): 102-10.
- [18] H2A: Hydrogen Analysis Production Models. National Renewable Energy Laboratory, 2018. <https://www.nrel.gov/hydrogen/h2a-productionmodels.html>.
- [19] Hydrogen Delivery Scenario Analysis Model (HDSAM). Argonne National Laboratory, 2006. <https://hdsam.es.anl.gov/index.php?content=hdsam>.

# Integrated Urban Energy Planning: A Case-Study Using Optimization

Jingbo GUO<sup>a1</sup>, Côme BISSUEL<sup>b1</sup>, Francois COURTOT<sup>b</sup>

<sup>a</sup>EDF (China) Holding Ltd., Tian Run Fortune Center A 807, 58, Dongzongbu Hu tong, Beijing 100005, China

<sup>b</sup>EDF R&D, 6 quai Watier, 78401 Chatou Catex, Paris, France

**Abstract.** This article describes an integrated energy planning optimization case-study. Starting from an integrated urban energy planning practice based on the urban planning information, an optimization approach is implemented to support decisions on suitable energy structures. Based on a use-case, energy demand, renewable energy resources, energy policy and energy prices are analyzed and set as inputs of the optimization. The results are energy structures minimizing the cost for two separated zones. Meanwhile, under different scenarios, in terms of renewable ratio targets and thermal storage, comparison is made for illustrating economy differences. The optimization mentioned in the article is modelled as a *Mixed-Integer Linear Programming* problem, which can search the optimal solution with high efficiency among the possible system designs.

**Keywords:** Integrated energy planning, optimization, renewable

## 1. Background

In China, traditional urban energy planning usually exists as several separated special energy planning (e.g., gas, electricity, heating) which are made by different institutes [1]. Consequences caused by these separated energy plannings are repetitive load prediction, oversized energy equipment, superfluous investment and inefficient energy operation [2]. Meanwhile, these “energy supply” oriented plans rarely consider energy structure, energy utilization mode, energy efficiency indicators as well as related environmental targets.

In recent years, increasingly urban green and low-carbon development popularizes distributed renewable, district energy, thermal storage, and local multi-energy system in the city. Integrated urban energy planning is emerging to better integrate all these energy aspects as a whole and efficient system. However, the diversity of energy demand, energy resources, technologies as well as local policies, increases complexity in the urban energy planning, which needs optimization tools to help find the best energy solution for cities.

---

<sup>1</sup>Corresponding authors: Jingbo Guo, Francois Courtot, EDF (China) Holding Ltd., Tian Run Fortune Center A 807, 58, Dongzongbu Hu tong, Beijing 100005, China. E-mail: jingbo.guo@edf.fr, francois.courtot@edf.fr; Côme Bissuel, EDF R&D, 6 quai Watier, 78401 Chatou Catex, Paris, France. E-mail: come.bissuel@edf.fr.

## 2. Introduction of optimization tools

The tool discussed in this paper is an optimization software for sizing a multi-energy system at district scale. It is developed under a *Mixed-Integer Linear Programming* framework to search an optimal energy solution, maximizing economy under various environmental and technical constraints in the city.

Inputs of the tool are the urban energy demand (electricity, heating, cooling) estimated based on the urban design, available potential of local renewable resources, energy tariff and related policies, as well as energy equipment information including efficiency, investment, maintenance cost, and service life.

Based on the annual estimated energy demand, we select several typical days and extreme days to represent the total demand and the extreme demand in a year. Local renewable potential will give the limitation when sizing an energy system. Energy tariff and equipment cost of each energy technology are part of the objective function.

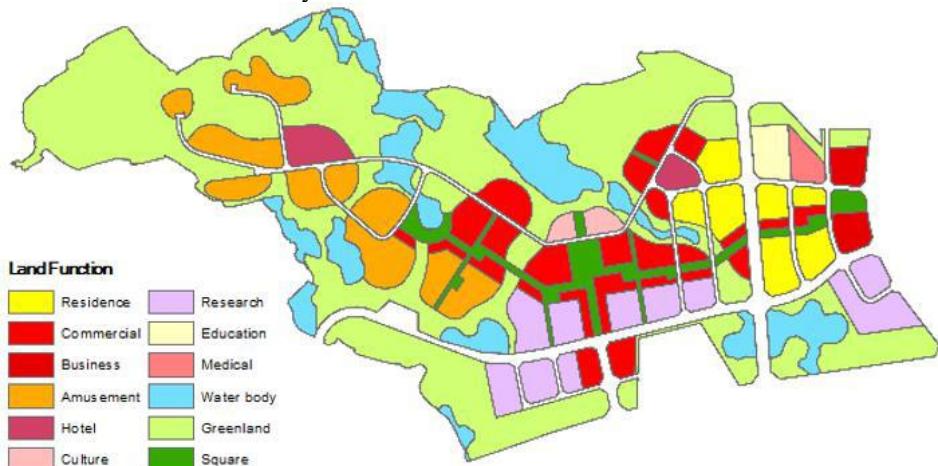
The output of the tool is an optimal multi-energy solution including the choice of energy technologies, the installed capacities, and the phasing scheme for a multi-phase deployment.

## 3. Case study

A use-case is made in a real city zone, demonstrating the approach in an integrated energy planning project.

### a. Introduction of the Zone

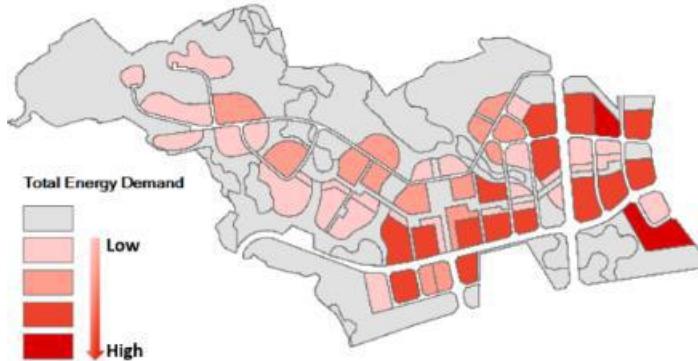
The use-case is located in a new urban area of a city in the middle of China, the optimization is applied during the urban design phase. The study zone covers an area of 1.85 km<sup>2</sup>, it is a multi-functions zone, and total floor area is 890,000m<sup>2</sup>, figure 1 shows the land function of the study zone.



**Figure 1.** Land function of the study zone.

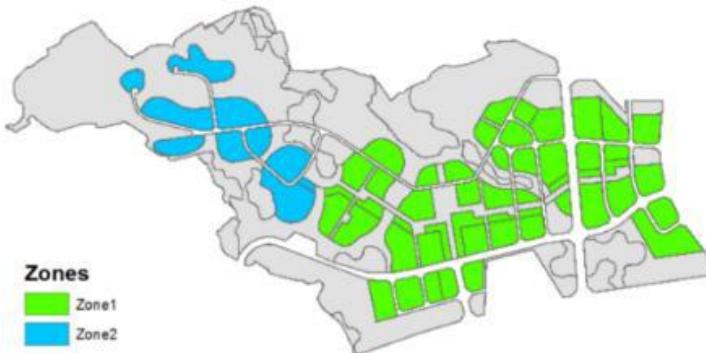
### b. Pre-study of Energy Condition

Based on land functions and floor area, a simulation is conducted to estimate energy demands (heating, cooling and electricity). By considering local climate, building types, envelope structure, operation mode, the annual hourly demand is estimated, which is one of the main input data for the optimization. The annual accumulated energy demand is then distributed to each plot, show in figure 2. The spatial distribution and the density of the energy demand provide valuable indications for positioning the energy systems.



**Figure 2.** Energy demand distribution.

The energy demand distribution shows that the east part has a larger and denser energy demand than the west; meanwhile, land functions in the east are more mixed which leads to diversified energy load features. Therefore, the study zone is divided into two parts, show in figure 3, the east part (zone1) will adopt centralized energy system and the west part (zone2) will use discrete energy solution at plot level. Floor area of these two zones is 800,000m<sup>2</sup> and 90,000m<sup>2</sup> respectively.



**Figure 3.** Classification of zones

By diagnosing the local renewable energy resources, both deep and shallow geothermal energy have application potential in the study zone. For deep geothermal energy, heat gradient in the study zone is 3.6°C per 100m depth, soil temperature could reach 100°C at 3km underground. Meanwhile, according to local geological data, it is suitable for geothermal drilling work. Thus, closed loop deep geothermal heat pump is considered as one heating solution. Local geological data also shows that abundant shallow soil-source geothermal energy exists in the study zone, so ground-coupled heat pump is another potential renewable technology.

Besides, roof solar and air-source heat pump are two other applicable energy resources in the zone. Meanwhile, due to the time-of-use electricity tariff in the city, thermal storage and power storage are regarded as potential solutions for increasing energy flexibility of local energy system.

By considering energy demand features and energy supply radius, maximum potential install capacity are detected for each technology and these values will be constraints when sizing energy system during optimization, table 1 shows the maximum install capacity of two zones.

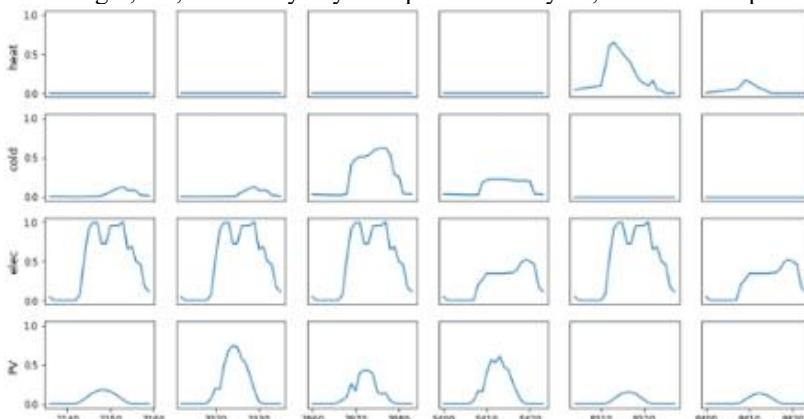
**Table 1.** Maximum install capacity of Zones.

Energy Technology	Zone1 Max Capacity [MW or MWH]	Zone2 Max Capacity [MW or MWH]
Gas Engine	no limitation	-
Absorption Chiller	no limitation	-
Electric Chiller	no limitation	no limitation
Gas Boiler	no limitation	no limitation
Deep Geothermal	15	1.5
Shallow Geothermal	1	1
Air-source heat pump	no limitation	no limitation
Roof PV	2.4	0.3
Elec Battery	no limitation	no limitation
Heat/Cold Storage	102/42	10.2/4.2

No limitation on a technology means that the energy source is abundant and would be sufficient to cover the total demand, in our use-case, these sources are the power grid and the natural gas network. Geothermal energy and thermal storage are mainly limited by available space and distance to the load center. Theoretically, roof PV could be installed on any free roof. By evaluating the radiation potential and the roof structure of the study zone, only parts of the roof are selected as good places for this technology, otherwise the installation cost will increase, and the unit electricity production will decrease.

### c. Optimization Study

The annual hourly energy demand and the energy technologies information are inputs of the optimization tool. To simplify the simulation work, 6 typical days and 4 extreme days are selected for each zone, show in figure 4. The selection of typical days and the associated weight, i.e., how many days it represents in a year, relies on an optimization.



**Figure 4.** Load curve of typical days of zone1.

Besides, local energy prices, cost of energy equipment, energy efficiency, and other related hypothesis are shown in Tables 2 and 3.

**Table 2.** Cost of energy equipment.

Energy Technology	CAPEX [rmb/MW or MWh or m <sup>3</sup> ]	Maintenance Cost per year [% of CAPEX]
Gas Engine	8,000,000	4%
Absorption Chiller	3,400,000	2%
Electric Chiller	1,600,000	2%
Gas Boiler	200,000	2%
Deep Geothermal	6,900,000	2%
Shallow Geothermal	4,600,000	2%
Air-source heat pump	2,550,000	2%
Roof PV	4,900,000	1%
Elec Battery	1,920,000	2%
Heat/Cold Storage	1,300	1%

**Table 3.** Current energy prices in the city.

Electricity price			
Price [rmb/kWh]	Peak	Flat	Valley
	1.1983	0.6657	0.3195
Peak period		10h-12h, 18h-22h	
Flat period		8h-10h, 12h-18h, 22h-24h	
Valley period		0h-8h	
Gas price			
Price [rmb/m <sup>3</sup> ]		3.094	

In this study, we also tested the impact of the renewable ratio on the economy of the whole energy system. Thus, in order to reach different renewable ratios, we extended the available roof surface for PV. The additional roofs have however lower solar radiation and small surface area which leads to higher unit cost and lower power production. Based on the local conditions, we defined three categories of roof for PV with the following maximum capacity and cost, shown in Table 4.

**Table 4.** Three levels of roof PV.

PV types	Zone1 Max Capacity [MW]	Zone2 Max Capacity [MW]	CAPEX [rmb/MW]
Low-cost PV	2.4	0.3	4,900,000
Middle cost PV	2.6	0.3	7,500,000
High-cost PV	7	1	9,000,000
Low-cost PV	2.4	0.3	4,900,000

Regarding calculation of the renewable energy ratio, different energy grades are taken into account. Meanwhile, primary energy efficiency is considered: only if the renewable technology has a higher primary energy efficiency than the business-as-usual solution (i.e., gas boiler for heating and electric chiller for cooling in our case study), such renewable technology is counted as an effective renewable energy [3].

The following equation is used for the renewable ratio calculation in our optimization:

$$\text{Renewable ratio} = \frac{\text{Energy produced by renewable} * \text{renewable share}}{\text{Total energy demand}} \quad (1)$$

The renewable share of each renewable energy is obtained according to the equivalence conversion coefficient of its end-use energy and its energy efficiency. The renewable shares are shown in Table 5.

**Table 5.** Parameter of renewable share.

Technology	Renewable share
Deep geothermal for heating	0.63
Shallow geothermal for heating	0.45
Shallow geothermal for cooling	0.21
ASHP for heating	0.27
ASHP for cooling	0
PV for electricity	2.2

#### d. General Study

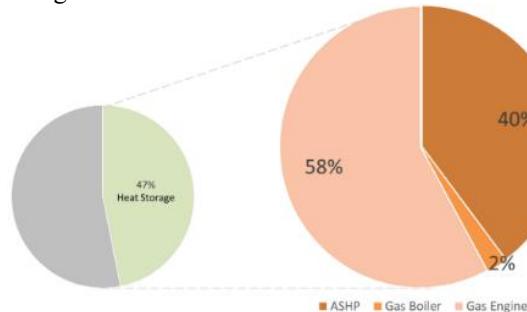
Based on all the inputs above, the optimization tool gives the technical solution with best economy for 20 years of service life.

Table 6 shows the optimal combination of technologies of zone1. Both heat and cold storages are installed at their maximum capacity. Traditional energy solutions are preferred in this case, both deep and shallow geothermal solution are abandoned by the tool. This shows these renewable solutions have less economic competitiveness in this zone. Total renewable ratio is 3.9%.

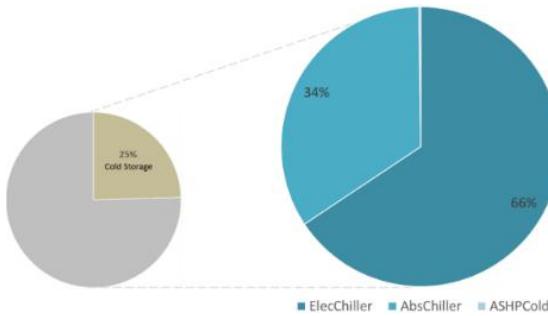
**Table 6.** Optimized installed capacity of zone1.

Installed heating capacity [MW or MWH]	Installed cooling capacity [MW or MWH]	Installed electricity capacity [MW or MWH]	
Gas Engine	7.14	Gas engine	7.14
ASHP	8.38	PV	2.4
Gas boiler	11.96		
Heat storage	102		
	Cold storage		

Looking at the energy production of each technology, gas engine and ASHP have produced the most part of heat energy, gas boiler only contributed to 2% of total heat production, mainly during peak hours. Almost half of the heat energy produced has been stored first, show in figure 5.

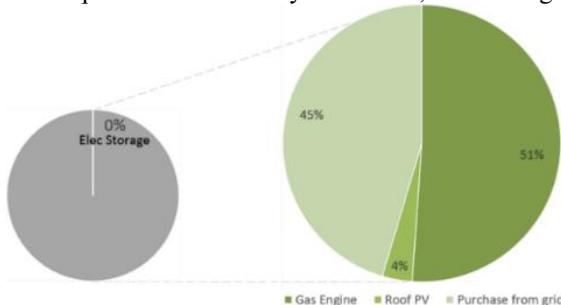
**Figure5.** Heat production structure of Zone1.

For cold energy production, electric chiller and absorption chiller undertake 66% and 34% of cold production respectively, ASHP only works for a few hours during high consumption days. A quarter of the cold energy has been stored, show in figure 6.



**Figure 6.** Cold production structure of Zone1.

Half of the electricity is produced by gas engines, roof PV contributes to another 4%, and the remaining part is purchased from grid. Due to the high cost of battery, all the electricity produced and purchased is directly consumed, show in figure 7.



**Figure 7.** Electricity production structure of Zone1.

In zone2, the most economical technology structure is shown in table 7. Compared to zone1, energy demand is too small to adopt gas engine which has a high unit cost. Shallow geothermal heat pump is identified as a good solution in zone2, and thermal storage is installed.

**Table 7.** Optimized installed capacity of zone2.

Installed heating capacity [MW or MWH]	Installed cooling capacity [MW or MWH]	Installed electricity capacity [MW or MWH]			
Shallow geothermal	0.92	Shallow geothermal	0.966	PV	0.3
ASHP	0.475	ASHP	0.57		
Gas boiler	0.045	Elec chiller	1.855		
Heat storage	10.2	Cold storage	4.2		

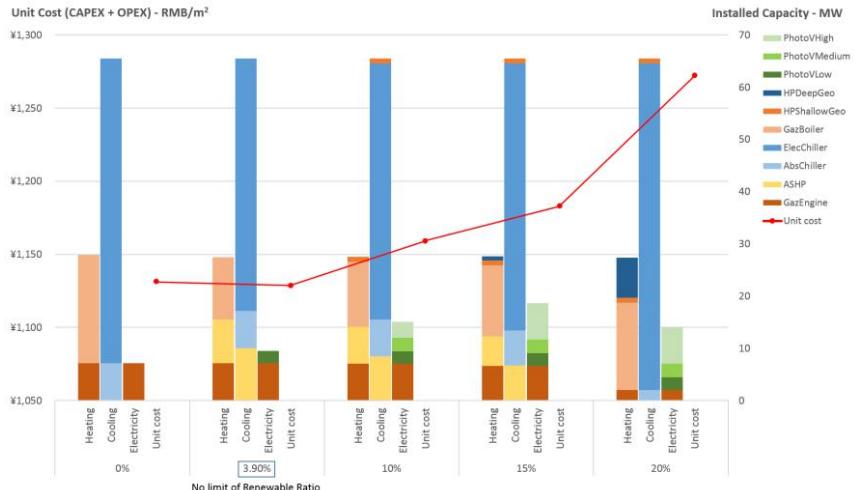
Regarding the energy production, shallow geothermal heat pumps generate 77% and 55% of the total heat and cold energy respectively; ASHP and electric chiller produces most of the rest; gas boiler and ASHP undertake the sharp peak demand. Nearly 96% of electricity is purchased from the grid, roof PV supply 4% of the total electricity demand. Under this solution, the renewable ratio of zone2 is 12.5%.

#### e. Renewable Orientated Results

Sometimes, city may have a mandatory or expected target of renewable ratio. In this case, economy will not be the only priority during optimization, but still the optimization tool could find a best solution under a given renewable ratio target constraint.

In this study, several energy solutions are given under different renewable ratio targets to analyze the impact on the economy of the whole system.

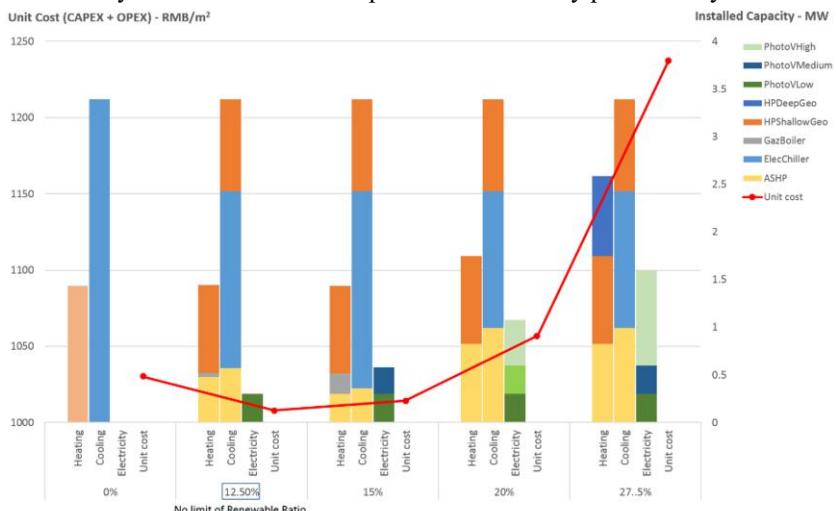
The optimal energy solutions obtained under different renewable ratio are shown in figure 8, giving the unit cost and energy structure comparison of each scenario during the whole service life in the zone1.



**Figure 8.** Unit cost and installed capacity of different renewable ratio in zone1.

In zone1, a certain part of renewable technology could improve the profit of the whole energy system; as the renewable ratio increases, the whole energy system will be more expensive, because high-cost PV and deep geothermal systems are built.

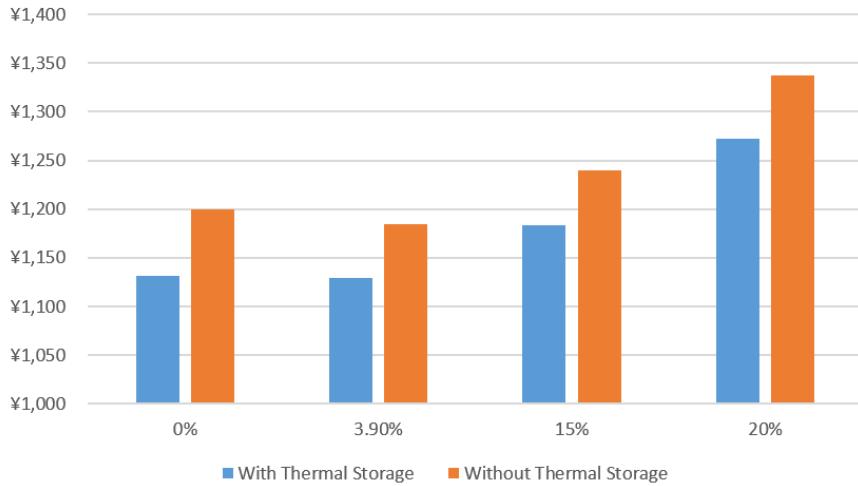
In zone2, the result shows the same trend, show in figure 9. When renewable ratio reaches 20% or a higher value, in order to satisfy such a high renewable ratio, the total installed capacity of heating is already more than the maximum demand, which means some equipment have to operate under low efficiency during certain period. Meanwhile, high-cost battery is installed to store superfluous electricity produced by roof PV.



**Figure 9.** Unit cost and installed capacity of different renewable ratio in zone2.

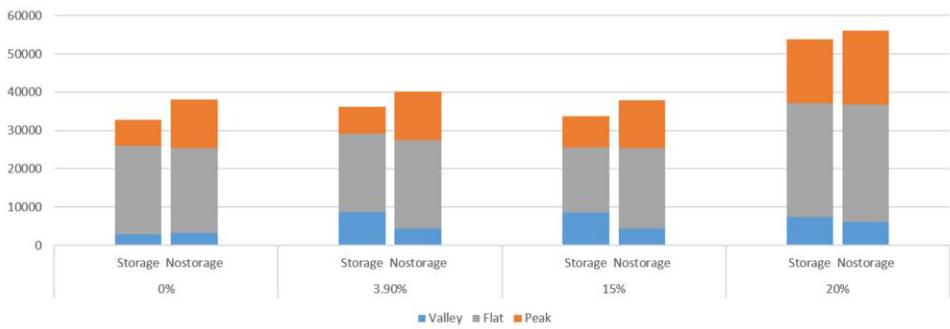
#### f. Impact of Thermal Storage in Multi-Energy System

In the study project, thermal storage reached its maximum installable capacity in all the energy solutions; thermal storage has a clear positive impact on economy of our multi-energy system. In zone1, by integrating thermal storage, the cost of energy system is reduced by around 5% in all renewable ratio target scenarios, show in figure 10.



**Figure 10.** Impact of thermal storage on economy.

When looking at the electricity consumption, as shown in figure 11, the thermal storage can shift parts of the peak electricity consumption to valley or flat periods of electricity tariff. Meanwhile, the thermal storage optimizes the operation of the different energy equipment by allowing the efficient equipment to operate for longer hours, increasing the global energy efficiency. Thus, the total electricity consumption decreases at around 10%.



**Figure 11.** Impact of thermal storage on electricity consumption.

## 4. Conclusion

Cities have diversified energy demand profiles, both in time and space, and diversified local resources. The green and low-carbon development backed up by proactive policies

unlocks the use of new technologies and new technologies combinations. This diversity of contexts and technical solutions greatly increases the complexity of urban energy planning. Optimization tools can bring a valuable support to guide the urban developers and designers to make sustainable energy choices from the early planning stage.

Until a certain renewable ratio, renewable energy is not only beneficial for the environment but also the most economical solution. Adapted choices in the urban planning phase can push further this point to increase the share of renewable energy that is economically accessible. The main levers are density of demand, mixity of functions, and localisation of consumption centres relative to the local resources. On the contrary, unreasonable renewable energy targets, not sufficiently considering local demand and resources, will increase the cost of energy system, and may damage the attractiveness of the new developed urban area.

According to the energy structure and local energy tariff, thermal storage could be a good solution in the multi-energy system to decrease both operation cost and investment of other equipment. This trend is expected to grow with the increasing need for integration of intermittent renewable energies.

## Acknowledgment

This paper is supported by the National Key Research and Development Program of China (Grant No.2018YFE0208400), the Science and Technology Project of State Grid Corporation of China (Key Technologies of Novel Integrated Energy System Considering Cross-border Interconnection)

## References

- [1] Guoxin, H., Zhihong, Y., Jinda Z., et al. Application challenges and prospects of urban integrated energy system planning. ARFEE 2019. 2020. E3S Web of Conference 143, 02017.
- [2] Weiding, L. Important stage of building energy efficiency: community energy planning. HV&AC. 2008. vol. 38, No. 3, 31-38.
- [3] Baoping, X., Wenlong, X. Study and discussion on assessment method of renewable energy utilization rate for new district planning. HV&AC. 2013. vol. 43, No. 10, 52-55.

# Study of Intelligent Search Engine of Energy Industry Based on BERT Semantic Model

Jiayang LI<sup>1</sup>, Hao LI, Ni YAN, Ziyun CHEN

Nari Group Corporation (State Grid Electric Power Research Institute), Nanjing Jiangsu 210061, China

**Abstract.** Based on a variety of heterogeneous data of energy enterprises, this paper provides an intelligent search method based on the Bert preprocessing model. This paper provides a method of intention recognition based on the combination of template matching and text classification, which involves machine learning, deep learning and other fields of artificial intelligence. First, the preprocessing model Bert is used to rewrite the natural language into a vector based on syntax. Then, the information extraction technology is used to extract the structured and machine understandable information from the problem to provide parameters for intention processing. The intention recognition technology used in this paper first uses template matching method, and then uses text classification method. Finally, the effects of various methods are compared through experimental examples.

**Keywords:** BERT model, word segmentation, search engine, deep learning

## 1. Preface

China's economic development has entered the "new normal", which is changing from high-speed growth to medium high-speed growth, and new changes have taken place in the energy industry. With the emergence and development of Internet of things, digitalization and online technology, the integrated dispatching construction of coal power energy enterprises has begun to transform to digitalization, networking and intelligence. The business of energy enterprises is complex, and there are a large number of structured, semi-structured data in the production and involving power, chemistry, railway, port, aerospace, petroleum and other professional fields. The data of energy enterprises has a large amount of data and complex structure, which is one of the difficulties in data analysis and retrieval.

Google team released a large-scale pre training language model [1] (Bidirectional Encoder Representation from Transformers, BERT) based on bidirectional transformer at the end of 2018 and this model refreshed the current optimal performance record of 11 NLP tasks by virtue of the pre trained fine turning model. The data of energy enterprises records a large number of business charact. At the same time, the text data also contains a

---

<sup>1</sup>Corresponding author: Jiayang Li, Nari Group Corporation (State Grid Electric Power Research Institute), Nanjing Jiangsu 210061, China; E-mail: 429901895@qq.com.

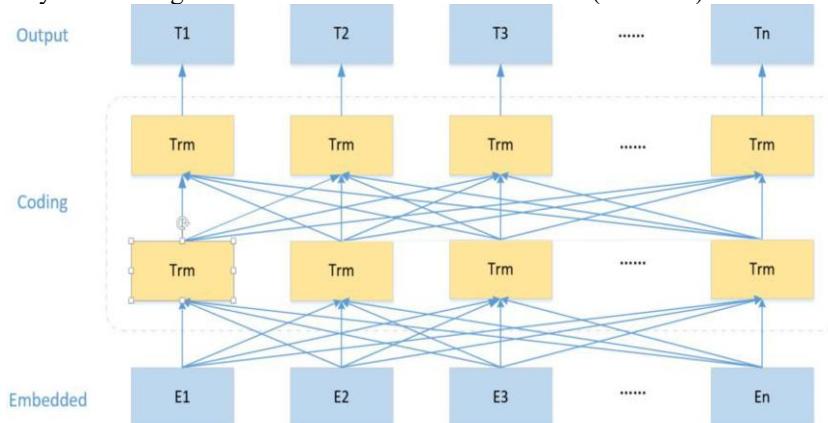
lot of industry terms, involving a wide range of fields, and has a huge amount. So, we need an effective intelligent search method to quickly retrieve the useful text data and find out the effective data information in the massive data.

According to the characteristics of the data in the energy industry, this paper focuses on the key technologies and core algorithms in the application of natural language processing, uses the pre training Bert model to analyze the content of questions and text data, mines the context rules of text data, and conducts secondary training through a large number of Chinese pre training models. The goal of this paper is to use natural language processing technology to transform questions suitable for the energy industry into machine language, provide reference suggestions for the application of natural language pre training model in the energy industry.

## 2. Model introduction

In this paper, the structure of the model is fine tuned based on the pre trained Bert line, so that it can be better applied to natural language processing and text classification tasks. We need to have a more comprehensive understanding of the structure and characteristics of the Bert model in order to make better use of the preprocessing model. From the perspective of professional technology, the Bert model does not use a new deep network or propose a new training algorithm, but improves and integrates the existing deep learning algorithm and natural language processing technology. Next, we introduce the Bert model.

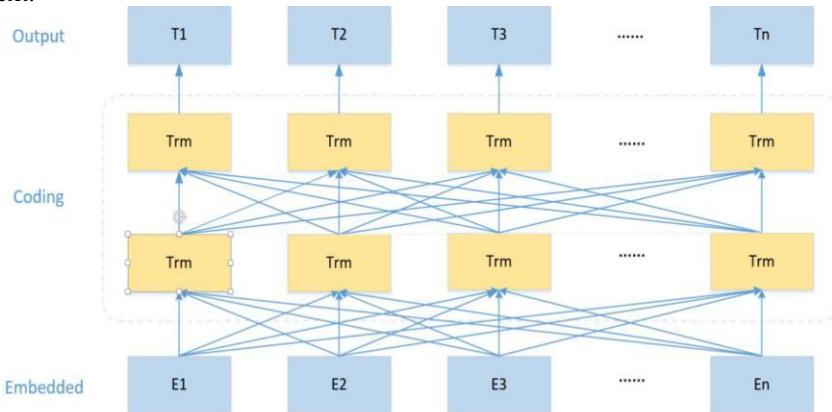
The full name of Bert model is Bidirectional Encoder Representation from Transformers. As shown in Figure 1, the Bert model [2] is mainly divided into two parts, the input layer and the bidirectional transformer coding layer. When processing a word, we can also consider the words before and after the word to get its meaning in the context. The attention mechanism of transformer has a good effect in feature extraction of words in context, and intuitively, considering the bidirectional encoding of context is better than only considering the unidirectional effect of the above (or below).



**Figure 1.** Basic structure of BERT model.

In natural language processing, word embedding is the basis of all researches. Using a vector to represent words and capture the semantics and relationships of words. A sentence is not a simple combination of some words, it must conform to human language

habits, and the language model is used to determine whether a sentence conforms to human habits, specifically, the word in front of a word (or context) in a given sentence, and then predict the probability of the word. Take an example of Word 2vec, which is proposed by Tomáš Mikolov in 2013, its feature is that all words are represented as low dimensional dense vectors, which can qualitatively measure the similarity between words in the word vector space. The core idea of word2vec is to hide a word from a sentence and train neural network to predict the hidden word. When the neural network training can successfully predict after a large number of corpus training, the word vector representation of the word can be settled through the neural network. However, with the increasing complexity of natural language processing tasks, word2vec can only return a word vector mixed with multiple semantics, which cannot solve the problem of polysemy. So, based on the problem that word2vec model cannot solve, researchers propose a new word embedding method (Embedding from Language Models, ELMo). In Elmo, the pre-trained model is no longer just a vector correspondence, but a trained model. When using Elmo model, input a sentence or a paragraph into the model, and the model will infer the corresponding word vector of each word according to the line text. One of the obvious advantages of this is that we can understand polysemous words in context.



**Figure 2.** Basic structure of Elmo model.

As shown in Figure 2 [3], each word in the sentence is encoded in three levels: In the coding layer, the first layer is the corresponding word coding in the forward LSTM, which mainly contains syntactic information; then there is the second layer of reverse LSTM, which contains semantic information. Finally, the embedded layer and the two coding layers are given a weight respectively, and the coding results are multiplied by their respective weights and summated to get the output of Elmo.

Combined with Figure 1 and Figure 2, it can be seen that there is no essential difference in the overall structure between Bert model and Elmo model. The fundamental difference is that the best model uses transformer encoder instead of LSTM encoding of Elmo model. The prototype of transformer includes two independent mechanisms, one encoder is responsible for receiving text as input, and the other is responsible for predicting the results of the task. BERT model is a multi-layer transformer encoder based on fine tuning, whose goal is to generate language models, so only the encoder part of transformer is required. By using transformer instead of Bi LSTM as encoder, Bert can have deeper layers and better parallelism.

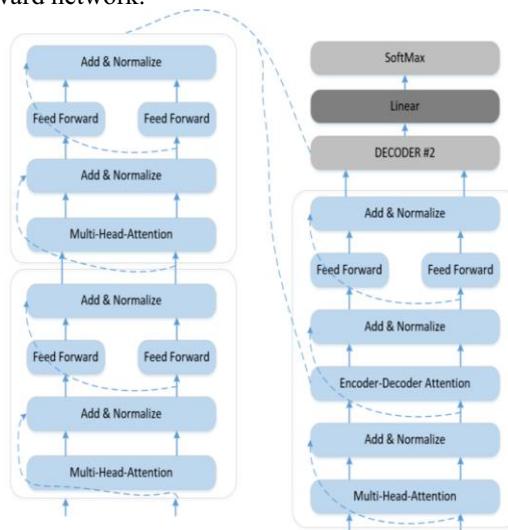
### 3. Construction of Intelligent Search Engine Model

#### 3.1. Transformer Encoder Based on Bert Model

Calculation steps of self-attention output:

- 1) For each word vector of the input encoder, multiply by three parameter matrices to create three vectors (query vector, key vector and value vector).
- 2) Calculate attention score: calculate the degree of attention a word places on words in other positions in the sentence, and calculate it through the dot product of the query vector corresponding to the word and the key vector of each word.
- 3) Divide each score by  $\sqrt{d_{key}}$ ,  $d_{key}$  is the length of the key vector.
- 4) These scores are normalized by a softmax.
- 5) Multiply the normalized score by each value vector.
- 6) Add the values in the previous step to obtain the output in the self attention layer.

In this paper, a fully connected network is added to the output of Bert and softmax is used as a classifier [4]. According to steps 2) - 6) above, the output of self attention can be obtained. As shown in Figure 3, the left part is the encoder, which is used to encode the source sequence; the right part is decoder, which is used to decode the output target sequence. As can be seen from the encoder on the left side of Figure 3, it is composed of N layers stacked together, and the structure of each layer is the same. Each layer consists of two sub layers, the first sub layer is composed of multi head attention sub layer and feed forward neural network sub layer. The second branch consists of a simple fully connected feedforward network.

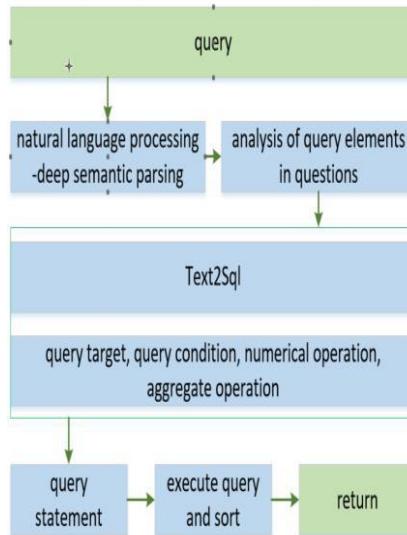


**Figure 3.** Transformer structure.

The decoder on the right is also made up of N-identical-layers. In addition to the two branch layers of the encoder, each layer adds a third branch layer coding decoding attention sub layer [5] (Encode-Decoder Attention), it is used to calculate the influence of output results in encoder on current results. In addition, the first multi head

self-attention sub layer is not exactly the same as the sub layer of encoder. Mask structure is added to the encoder to deal with the words shielded in the training stage. The decoder also performs residual connection and layer normalization.

### 3.2. Intelligent Search Engine Architecture



**Figure 4.** Search engine process.

The processing flow of search engine is described. As shown in Figure 4, according to the user's question, the intelligent search engine uses natural language processing technology to analyze the element information in the question, recognize and classify the text [6]. The natural language is transformed into SQL statements, and the returned results are sorted. Current search engines tend to pay attention to generality, but invest less in the acquisition and analysis of user query requests. Usually search engine only supports keyword search and logical operation on this basis [7]. However, the search engine proposed in this paper has made the following two improvements in data preprocessing:

(1) Using natural language processing technology, Extract the effective components of query conditions, including lexical information and its logical relationship, to prepare for intention recognition.

(2) Through the study of superior words and inferior words, according to the user-defined dictionary to get synonyms, synonyms and Related words of keywords.

The above analysis is mainly for vocabulary. When you enter a question with complex logical relationship, it is difficult to combine the search terms with keywords. So, the use of natural language processing technology is the ultimate and fundamental way to achieve the accurate description of the search and ensure the search relevance and similarity. Next, let's introduce the detailed steps of intelligent search engine in this paper:

- 1) Data input: A user's question is described as a piece of data, so that the data set to be trained is  $Zy = \{Zy_1, Zy_2, Zy_3, \dots, Zy_n\}$ ;

- 2) Chinese word segmentation: This paper uses the CRF word segmentation algorithm [8], at the same time, it updates the energy industry based terminology and dictionary to segment the question text  $A_{ij} = \{A_{11}, A_{12}, A_{13}, \dots, A_{jn}\}$ ; one question corresponds to the formula I, each of which is  $j$ ; Here, the conditional probability calculation formula represented by CRF model is:

$P(Y|x; \theta) = \frac{1}{Z_x(\theta)} \exp\{\sum_{i=1}^k \theta_i f_i(D_i)\}$  where, the denominator is the partition function, and the expression is:  $Z_x(\theta) = \sum_Y \exp\{\sum_{i=1}^k \theta_i f_i(D_i)\}$ , when using the optimization method to train the parameters of CRF model, the main task is to calculate the cost and grad expressions of the model.

- 3) Data preprocessing: Remove redundant words in the text, such as non-functional auxiliary words, adverbs, prepositions, conjunctions and so on.
- 4) Text representation: Using Bert pretreatment model, each word after text segmentation is transformed into a vectorized value, Using the vectorized value to find the optimal matching result;
- 5) Feature construction: construct the corresponding feature thesaurus matrix to complete feature extraction, which can effectively reduce the dimension of feature space and improve the operation efficiency:

$$\sum A_{IJ} = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1j} \\ A_{21} & A_{22} & \dots & A_{2j} \\ A_{i1} & A_{i2} & \dots & A_{ij} \end{bmatrix}$$

6) Model training: in this method, three models need to be trained, that is, the classification model of predicting table names; The classification model of predicting column names and the model of converting to SQL statements. The feature matrix is used as the input of the algorithm model, of which 70% is the training set and 30% is the test set;

7) Evaluation process: This paper uses the accuracy, recall, precision and other four indicators to evaluate the results. In fact, the accuracy rate is for our prediction results, which indicates how many of the predicted positive samples are really positive samples. Then there are two possibilities for positive prediction. One is to predict the positive class as the positive class (TP), and the other is to predict the negative class as the positive class (FP). Recall rate indicates how many positive examples in the sample are predicted correctly. The accuracy reflects the closeness of the measurement result to the true value.

■accuracy:

$$ACC = \frac{TP+TN}{P+N} = \frac{\text{number of true predictions} + \text{number of false predictions}}{\text{number of correct predictions} + \text{number of errors}}$$

■recall:

$$R = \frac{TP}{TP+FN} = \frac{\text{Number of true prediction}}{\text{number of true prediction} + \text{number of false prediction errors}}$$

■precision:

$$P = \frac{TP}{TP+FP} = \frac{\text{Number of true prediction}}{\text{number of true prediction} + \text{number of false prediction}}$$

#### 4. Experimental Data and Analysis

In the production and operation practice of the national energy group, priority is given to ensuring the continuity of coal production and coal transportation process and maximizing the utilization of production capacity, so as to "promote" the efficient operation of the industrial chain. In the process of coal production and transportation, multi link manual coordination is carried out to dynamically balance the coal demand and ensure the coal supply of internal and external power plants. The overall industrial chain operation belongs to "production promotion", Production and operation mode of "two-way balance between supply and demand". The training data of this paper are the data of electric power, coal, integration, sales, chemistry, railway, port, shipping, oil and other modules in the energy industry. The data of each category are shown in Table 1.

**Table 1.** samples.

field	category	samples
coal	coal_t	12003
power	pow_t	11908
railway	rail_t	9872
port	port_t	8794
ship	ship_t	8976
chemical	che_t	9321
sale	sail_t	10023
integrated	uni_t	6532

The model training is carried out on Tesla V100 GPU, and the training parameters are shown in Table 2.

**Table 2.** training parameters.

parameter	category
network layers	24
Hidden layers	768
Self-Attention-Head	12
Parameters	340M
Activation function	GLUE
epochs	50

The learning and training of the algorithm model adopts the method of repeated sampling with putting back to construct the training set and the test set, in which the training set and the test set are divided according to the proportion of 70% and 30%. In the training process of this paper, in order to train a deep two-way language vector, we need to cover some words in the sentence. The shielding strategy is 80% probability to keep the word unchanged, 10% probability to replace with a random word, and 10% probability to replace with mask token.

Three models will be trained in this paper. The first model is the classification model to determine the table name; The second model is to determine the column name; The third model is the model file that generates SQL.

In order to reasonably evaluate the effect of intelligent search semantic analysis, we focus on the accuracy, recall and accuracy of the model. In addition, in order to compare the effect of BERT model and traditional machine learning classification model, this paper selects several typical machine learning algorithms for comparative experiments. In this paper, LSTM long-term memory network and RNN recurrent neural network are used for experimental comparison. The results are shown in Table 3:

**Table 3.** experimental result.

Algorithm model	Accuracy	recall	precision;	time
Pre-training model BERT	0.9551	0.899	0.832	392s
LSTM	0.921	0.833	0.765	345s
RNN	0.889	0.863	0.835	492s

It can be seen from table 3 that the training time of LSTM using long-term and short-term memory network is the least, but the accuracy, recall and precision indicators are not as good as the pre training model Bert. The pre training model Bert has good performance in accuracy, recall, accuracy and other indicators. The pre training model Bert can bring the text features of domain training to the scene of search engine, and can significantly improve the accuracy.

## 5. Conclusion

In the intelligent search engine based on energy system data, by introducing the Bert model based on pre training, and comprehensively comparing the pre-processing model with the commonly used machine learning and deep learning models, by comprehensively comparing the accuracy, recall, accuracy and time-consuming of other algorithm models, and training the model suitable for energy system, an optimal scheme of intelligent search on national energy system is given.

At present, in addition to the more mature Bert model, there are a series of more complex preprocessing models such as mass and xlnet. It can be expected that with the development of artificial intelligence, natural language processing and other technologies, the intelligent search engine based on the national energy system business will usher in great changes.

## Reference

- [1] Hu CT, Qin JK, Chen JM, Zhang L. Research on application of public opinion classification based on Bert model. *Network Security Technology and Application*, 2019 Nov; 2019(11):41-44.
- [2] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [J] 2018.
- [3] Tomas M, Kai C, Greg C, and Jeffrey D. Efficient estimation of word representations in vector space. In Proceedings of Workshop at ICLR, 2013.
- [4] Ren Y. Power grid a classification model of power work orders texts based on pre-trained BERT model. *Yunnan Electric Power*. 2020; 048(001):1-7,11.
- [5] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences [J]. *Eprint Arxiv*, 2014, 1.
- [6] Liu XH, Chen WS, Ren QN, News text classification based on attention cnlstm network model. *Computer Engineering*. 2019 Jun; 56(14):7.
- [7] Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need [J]. 2017.
- [8] Zhang K. Application and development of transmissionline inspection system based on multi-rotor UAVs [J]. *Modern Industrial Economy & Informationization*, 2013.

# Game Analysis on Profits of Water Diversion Project Supply Chain Under the Overall Loss of Supply Chain in Undeveloped Areas

Jinren LUO, Jiayin LI<sup>1</sup>, Qinyao FU, Yihu ZHAO, Shan LIU

*Institute of Economics and Management, Lanzhou Jiaotong University, Lanzhou, Gansu, 730070, China*

**Abstract.** This paper compares and analyzes the effects of different amounts of government subsidies and subsidy patterns on water supply, pricing, and profit distribution, and takes water diversion project supply chain composed of water transfer company and the water work as the study object, using financial accounting method innovatively, establishing non-cooperative and cooperative game models of two under government subsidies. The results show that as the number of subsidies increased, the optimal amount of water supply and the profits increased, but the price of water work decreased. When the amount of subsidy stays same, the price of water plant also stays same. However, the price of the water transfer company is constantly changing, the higher the proportion of subsidies it received, the lower the price. In terms of data, the authors take a practical example - the Tao River Water Diversion Project to analyze. This paper's limitation is that the conclusion is based on a single water transfer company and a single water plant as the research object. However, the actual situation of the water transfer project is that a water transfer company faces a complex supply chain network formed by many water plants and farmers' water users' associations.

**Keywords.** The supply chain of water diversion project in undeveloped areas, government subsidies, non-cooperative game, cooperative game, Shapley Value

## 1. Introduction

Water diversion project is an effective measure to solve the problem of highly uneven distribution of water resources in the world. With the development of various national undertakings, cross-regional water transfer projects are increasing in China. South to North Water Division Project, Yellow River to Qinghai Water Division Project, and Tao River Water Division Project are based on multi-objective water supply, serving urban industry and considering agriculture and environment, which are both operating and public welfare. In terms of the situation that the affordability of water price in some underdeveloped areas cannot even make up for the project operation cost, it is necessary to consider the government subsidy under the premise of considering the public welfare

---

<sup>1</sup> Corresponding Author: Jiayin Li, Institute of Economics and Management, Lanzhou Jiaotong University, Lanzhou, Gansu, 730070, China; E-mail: 1948858040@qq.com

This work is supported by National Natural Science Foundation of China [71463036]: Benefit Balance Mechanism of Stakeholders with Multi-objective Trans-water Project Between Less Developed Regions.

of the project and the water price born by users. How can government subsidies be introduced to balance the interests of enterprises in the water transfer project supply chain in underdeveloped areas to maintain sustainable water supply has become an important research topic.

Domestic and foreign scholars have obtained numerous research results on the issue of government subsidies in the supply chain. Their research focuses mainly on the effects of subsidies, subsidy objects, and subsidy weight, and the game is the major method. The effects of subsidies and subsidy objects are studied mostly from the qualitative perspective. T.G. Mallory [1] assessed the causation and effectiveness of Chinese government fishery subsidies and the effects of government subsidies on fisheries. L.H. He and L.Y. Chen [2] studied the incentive effects of four different government subsidy policies in the green construction market. D. McQuestin and M. Noguchi [3] analyzed the effectiveness of government subsidies in the emergency waiting of local health network systems. Frye and Shleifer(1997) pointed out that in the transitional economy, subsidies are the most direct means for the government to play a ‘helping hand’ role. [4] R.C. Hu et al. [5] studied the application effect of government subsidies at the provincial and industrial levels. H.J. Wang et al. [6] studied the government’s preference to subsidize enterprises with serious losses (poor behavior) or enterprises with light losses (emergency behavior) and their economic consequence from the perspective of property rights nature, regional government intervention degree, and industry competition degree. D.M. Kong and T.Sh. Li [7] investigated the effects of government subsidies on enterprises with different property rights from the perspective of business performance and social responsibility. Ch.S. Wu et al. [8] studied the effects of government subsidies on business performance from the political connection perspective. Q. Geng and R.X. Hu [9] believed that the probability and degree of enterprises obtaining subsidies are affected by their endowment and nature. Generally speaking, Chinese government subsidies have an obvious state-owned preference, scale preference, export preference, and industry tendency.

Many scholars use game theory to study the supply chain weight. Q. Wu et al. [10] used the core method of the cooperative game and established the mix-integer linear programming model to achieve fair benefit distribution among participants in the distributed energy network. M. Babaei et al. [11] proposed that government subsidies to investors and control the number of intermediaries can promote investors to earn profits, which is supported by Stackelberg game theory. Q.H. Zhu and Y.J. Dou [12] established a three-stage game model that considers the green degree of products and government subsidies in the green supply chain. H.X. Lu [13] constructed a Stackelberg decentralized decision model and centralized decision model of the supply chain under government subsidies and analyzed the effects of the changes in government subsidies on pricing decisions. Y.D. Li et al. [14] constructed a game model in which the government subsidizes the cooperative emission reduction investment of the low-carbon supply chain under three different game relationships: Nash game, Stackelberg game, and centralized decision-making of supply chain. They also analyzed the optimal emission reduction cost of the supply chain and the optimal subsidy rate of the government. X.L. Zhang and J.J. Wang [15] used the Shapley value method to study the impact of government subsidies on supply chain decision-making in the new energy vehicle supply chain.

In the existing research results, many analyses of the game relationship between water transfer companies and tap water plants can be found. For example, H.M. Wang et al. [16] and Zh.S. Chen [17] established some interesting game models and analyzed the relationship between water transfer companies and tap water plants in the water supply

chain. Some scholars also analyzed the relationship between the supply chain of water diversion projects and government subsidies from the theoretical perspective. L.Zh.Wu et al. [18] explored the multi-objective reservoir optimal operation model and carried out inter-basin water transfer under the consideration of the benefits contradiction between environmental and economic. X.Zh. Wang and Y.R. Wang [19] considered the public welfare of the South to North Water Division Project and the affordability of water price users, analyzed the reasonable water price composition of the project, and studied the water price subsidy policy. J.R.Luo [20] calculated the affordable water price of different water used in the Tao River Division Project and put forward the suggestion of government subsidy for agricultural water with low bearing capacity.

Academic circles have demonstrated the importance of government subsidies for economic development and studied government subsidies objects, weight in the supply chain by the method of game. These studies have some guiding significance for the current interest coordination management of water supply project's supply chain, but there are still some shortcomings: most studies are limited to the theoretical models. Because the model is not integrated with the specific financial accounting information, the parameters of the models are difficult to determine, which leads to the poor guidance of the established mathematical model. The research on the subsidy of inter-basin water transfer projects is not mature, and limited research has focused on the subsidy of the water resource supply chain.

Given the above considerations, this paper intends to use the financial accounting method to establish a two-level supply chain interest game model composed of water transfer companies and water plants (including farmers' water users' associations) participated by the government, and analyze the effects of different government subsidy amount and subsidy mode on water supply quantity, pricing, and profit distribution of water transfer companies and waterworks. The data of the Tao River Division Project will be used to analyze the case.

## 2. Description of the Basic Concept

### 2.1. Method of Shapley Value

In 1953, Shapley proposed the Shapley value method to solve the problem of profit distribution in multi-person cooperative games. When  $n$  individuals are engaged in economic activity, each form of cooperation among several of them will produce certain benefits. When the interest activities between people are non-antagonistic, the increase in the number of people in cooperation will not cause a decrease in income, then the cooperation of all  $n$  individuals will bring the maximum benefit. In other words, when  $n$  individuals form an alliance, they will obtain the maximum benefit. The Shapley value method is a scheme to redistribute the maximum income among participants. Its connotation is as follows: Let set  $I : \{1, 2, \dots, n\}$  if any subset  $X$  (representing any combination in the set of  $n$  people, also known as a coalition) of  $I$  corresponds to a real-valued function,  $v(X)$  satisfying the

$$v(\emptyset) = 0 \quad (1)$$

$$v(X_1 \cup X_2) \geq v(X_1) + v(X_2), X_1 \cap X_2 = \emptyset \quad (2)$$

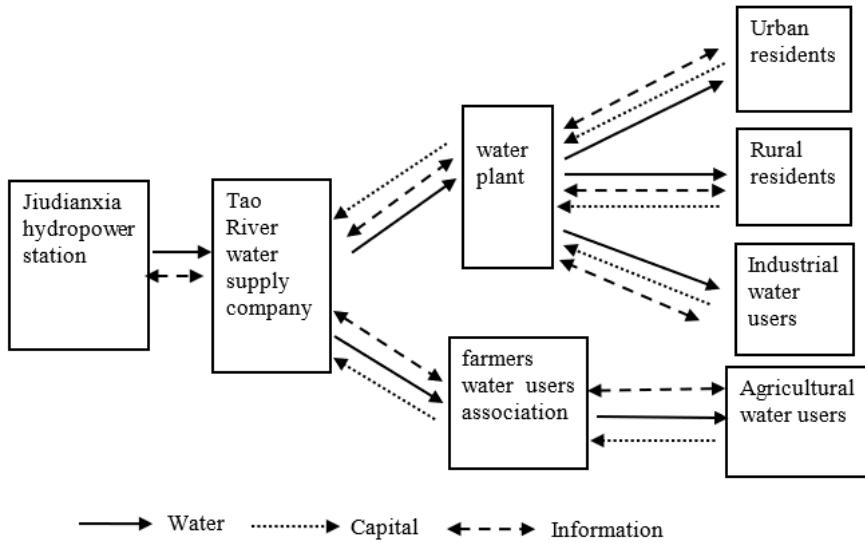
Then the  $[I, \nu]$  is called  $n$  people cooperation countermeasures, where  $\nu$  is the characteristic function of cooperative countermeasures. In this paper, we use  $\varphi_i(\nu)$  to denote  $i$  members' income allocated by  $I$  members from the maximum profit  $\nu(I)$  of cooperation. The set of the allocation of  $n$ -person cooperative countermeasures is  $\Psi(\nu) = (\varphi_1(\nu), \varphi_2(\nu), \dots, \varphi_n(\nu))$ . The success of cooperation must satisfy the following conditions:

$$\sum_{i=1}^n \varphi_i(\nu) = \nu(I), \text{ and } \varphi_i(\nu) \geq \nu(i), i = 1, 2, \dots, n \quad (3)$$

The profit distribution of each partner under cooperation  $I$  is called the Shapley value. The expression of Shapley value is as follows:

$$\varphi_i(\nu) = \sum_{X \subseteq N} \frac{(|X|-1)!(n-|X|)!}{n!} [\nu(X) - \nu(X - \{i\})], \forall i \in N \quad (4)$$

This expression  $\varphi_i(\nu)$  represents the Shapley value of the  $i$ th member in the supply chain.  $|X|$  is the number of members in the  $X$  subset,  $n$  is the total number of members in the supply chain,  $\nu(X)$  is the profit value of the  $X$  supply chain subset, and  $\nu(X - \{i\})$  is the profit value not included in the  $X$  supply chain subset.



**Figure 1.** Water Resources Supply Chain Diagram of Tao River Diversion Project.

## 2.2. Water Resources Supply Chain Diagram of Water Transfer Project

Water Division Project refers to the water diversion system composed of hydro-junction, net pipe, and pumping station. Taking the Taohe River Division Project as an example, the water resources supply chain of the water diversion project is a complex network system [21]. In this system, the water flows down from the Jiudianxia hydropower

station to the users in the receiving area, and the fund flows upward from the users in the receiving area to the water division company of Tao River. The supply and demand information flows in two directions in the entire chain. The schematic diagram is shown in Figure 1.

### 3. Game Analysis of Supply Chain Benefits Under the Premise of The Overall Loss of Water Transfer Project Supply Chain

#### 3.1. Model Analysis

Limited by the level of economic development, the affordability of water prices in underdeveloped areas is generally low, and the affordability of different water use categories also differ. Taking the Tao River Diversion Project as an example, the bearing capacity of domestic water of rural and urban areas and industrial water is greater than the full cost of the supply chain. While the bearing of agricultural water is far lower than the cost of the supply chain, it cannot make up for the operation cost of the water transfer company. [21] In the view of the later situation, this paper explores how the losses can be shared between water transfer companies and waterworks, and how it can be subsidized by the government to maintain the normal operation of the water diversion supply chain.

The total cost of the water transfer project supply chain can be divided into two parts: fixed cost and variable cost. Because fixed cost (formed mainly by fixed assets invested by state) has become a sunk cost, it will happen even no water supply can be found. The variable cost will be put into the case of water supply. For the node enterprises, whether water supply or waste use depends on whether their variable costs can be compensated. If  $V_m$  and  $V_r$  are used to represent the unit variable cost of the water transfer company and water plant, respectively,  $Q$  is the market demand for water supply,  $P$  is the water price of the water plant. When the fixed cost is not considered  $\pi_m$  and  $\pi_r$  is used to represent the income of the water transfer company and the water plant respectively. The income matrix of both parties is shown in Table 1.

**Table 1.** Income matrix 1 of water transfer companies and waterworks.

		Water works	
		strategy	water use
Water transfer company	water supply	$\pi_m, \pi_r$	$-V_m Q, 0$
	no water supply	$0, -V_r Q$	$0, 0$

In the case of  $P < V_m + V_r$ ,  $\pi_m + \pi_r = P Q - (V_m + V_r) Q < 0$ , that is, if the water price of the waterworks is lower than the sum of the variable costs of the water diversion company and water plant, regardless of the rules that the two sides use to distribute profits, there must be at least one sides' income being negative, and the party with a negative income will inevitably choose the strategy of no water supply or no water use. Thus, the best strategy of the other party at this time will be no water use or no water supply. Therefore, the optimal strategy combination for both sides is (no water supply, no water use), and the revenue is 0, which is a typical Nash equilibrium.

The water transfer projects with public welfare in underdeveloped areas requires that government subsidies should be considered, which makes water division companies and water plants willing to choose the strategic combination (water supply, water use). Without considering the fixed cost, if the subsidy is less than the loss of the variable cost of the supply chain, it means at least one party's income is negative after the subsidy, and two sides will choose the (no water supply, no water use) strategy combination. However, previous research on the subsidies shows that more subsidies are not better. On the contrary, it may reduce the operating efficiency of the supply chain. The best strategy is that the subsidy amount is exactly equal to the total variable cost loss, and the variable cost loss of the water transfer company and water plant is just compensated. The income matrix of both parties is shown in Table 2.

**Table 2.** Income matrix 2 of water transfer companies and waterworks.

		Water works	
		strategy	water use
Water transfer company	water supply	0,0	$-V_m Q, 0$
	no water supply	$0, -V_r Q$	0,0

At this time, although many equilibrium solutions can be obtained in theory, considering the role of government and cooperation of social responsibility, we can see that the optimal strategy combination for both sides is (water supply, water use), which is beneficial to the three parties.

### 3.2. Model Hypothesis and Variable Description

#### 3.2.1. Model Hypothesis

First, the game participants include the government, water transfer companies, and waterworks (including farmers' water users' association). Because the government only plays a regulatory role, the specific game is only between the water transfer company and the water plant, both of which are bounded, rational economic men.

Second, the government's strategy set is {all subsidies, subsidies only to water transfer companies, subsidies only to waterworks, no subsidies}, the strategy set of water transfer companies is {water supply, no water supply}, and the strategy set of waterworks is {water use, no water use}. After the government subsidies are added, the decision-making order is the government → water transfer company → water plant.

Third, no other source of water for the waterworks other than the water diversion company can be found.

Fourth, the demand for tap water is a decreasing function of the retail price, and the demand function is as follows:

$$Q = Q_0 - \alpha P \quad (5)$$

In other words,  $Q_0$  represents the maximum market demand and  $\alpha$  is the sensitivity coefficient of sales volume to the price of tap water,  $\alpha > 0$ ,  $Q > 0$ ,  $Q$  is equal to the water production, water supply of the water transfer company, and the water plant.

### 3.2.2. Model solution

This paper calculated the incomplete cost of water transfer company and waterworks according to the method of financial accounting after deducting the depreciation of fixed assets and dividing them into unit incomplete cost according to the volume of water.

The unit cost of water transfer company  $C_m = (\text{operating cost } C_{m1} + \text{interest on fluid capital } C_{m2})/Q$ , where operating cost  $C_{m1} = \text{engineering maintenance cost } C_{m11} + \text{wage welfare fee } C_{m12} + \text{project management fee } C_{m13} + \text{power cost } C_{m14} + \text{other expenses } C_{m15}$ .

The unit water distribution cost of water plant  $C_r = (\text{main business cost } C_{r1} + \text{period cost } C_{r2} + \text{other expenses } C_{r3})/Q$ , in which period cost  $C_{r2} = \text{management expense } C_{r21} + \text{operation expense } C_{r22} + \text{financial expense } C_{r23}$ .

The water price of the water transfer company is  $P_m$ , and that of the water plant is  $P$ .

The total amount of government subsidies is  $S = S_m + S_r$ , that of water transfer companies is  $S_r = (1 - b)S$ ,  $0 \leq b \leq 1$ .  $b=1$  is to subsidize the water transfer company alone,  $b=0$  is to subsidize the water company only,  $0 < b < 1$  is to subsidize the water transfer company and the water supply company.

### 3.3. Model Construction and Solution

#### 3.3.1. Model Construction

The profit function of a water transfer company is as follows:

$$\Pi_m = (P_m + bS - C_m) * (Q_0 - \alpha P) \quad (6)$$

The profit function of waterworks is as follows:

$$\Pi_r = [P + (1 - b)S - P_m - C_r] * (Q_0 - \alpha P) \quad (7)$$

The gross profit function is as follows:

$$\Pi = (P + S - C_m - C_r) * (Q_0 - \alpha P) \quad (8)$$

#### 3.3.2. Model Solution

(1) An analysis of the non-cooperative game between the water transfer company and the water plant

In daily life, when the water transfer company determines the water price  $P_m$ , the water company first determines the market price  $P$  according to the price of the water transfer company and other factors, and the two parties form a Stackelberg game relationship. Therefore, the equilibrium solution of the game can be obtained according to the reverse induction method [22].

Formula (7) derives the first derivative of  $P$ ,  $\frac{\partial \Pi_r}{\partial P} = Q_0 - 2\alpha P + \alpha P_m + \alpha C_r - \alpha(1 - b)S$ . Let  $\frac{\partial \Pi_r}{\partial P} = 0$ . The optimal water price for a water company in a non-cooperative game is obtained as follows:

$$P^{\text{non cooperation}} = \frac{Q_0 + \alpha P_m + \alpha C_r - \alpha(1 - b)S}{2\alpha} \quad (9)$$

By substituting Formula (9) into Formula (6), the profit function of the water diversion company is obtained as follows:

$$\Pi_m^{non\ cooperation} = \frac{1}{2}[Q_0 - \alpha P_m - \alpha C_r + \alpha(1-b)S](P_m + bS - C_m) \quad (10)$$

Formula (10) obtains the first derivative of  $P_m$ ,  $\frac{\partial \Pi_m^{non\ cooperation}}{\partial P_m} = \frac{1}{2}(Q_0 + \alpha C_m + \alpha S - 2\alpha bS - 2\alpha P_m - \alpha C_r)$ . Let  $\frac{\partial \Pi_m^{non\ cooperation}}{\partial P_m} = 0$ , the optimal selling price of the water transfer company is obtained as follows:

$$P_m^{non\ cooperation} = \frac{Q_0 + \alpha C_m + \alpha(1-2b)S - \alpha C_r}{2\alpha} \quad (11)$$

Substituting Equation (11) into Equation (9), the optimal water price of a water company is as follows:

$$P^{non\ cooperation} = \frac{3Q_0 + \alpha C_m + \alpha C_r - \alpha S}{4\alpha} \quad (12)$$

By substituting Equation (12) into Equation (5), the market demand for tap water is obtained as follows:

$$Q^{non\ cooperation} = \frac{1}{4}[Q_0 + \alpha S - \alpha C_m - \alpha C_r] \quad (13)$$

By substituting Formulas (11), (12), and (13) into formulas (6), (7), and (8), the total profits of water diversion company, water plant, and supply chain are respectively as follows:

$$\Pi_m^{non\ cooperation} = \frac{[Q_0 + \alpha S - \alpha C_m - \alpha C_r]^2}{8\alpha} \quad (14)$$

$$\Pi_r^{non\ cooperation} = \frac{[Q_0 + \alpha S - \alpha C_m - \alpha C_r]^2}{16\alpha} \quad (15)$$

$$\Pi^{non\ cooperation} = \frac{3(Q_0 + \alpha S - \alpha C_m - \alpha C_r)^2}{16\alpha} \quad (16)$$

The derivative of Equation (12) to S is obtained  $\frac{\partial P^{non\ cooperation}}{\partial S} = -\frac{1}{4} < 0$ . That is, the price of waterworks will decrease with the increase in government subsidies. Formula (13) derives S and obtains  $\frac{\partial Q^{non\ cooperation}}{\partial S} = \frac{\alpha}{4} > 0$ , that is, the demand for tap water will increase with the increase in government subsidies.

(2) An analysis of the cooperative game between the water transfer company and the water plant

Formula (8) derives the first derivative of price  $p$ ,  $\frac{\partial \Pi}{\partial p} = Q_0 + \alpha C_m + \alpha C_r - 2\alpha P - \alpha S$ . Let  $\frac{\partial \Pi}{\partial P} = 0$ , the optimal water price for a water company in the case of cooperation is as follows:

$$P^{cooperation} = \frac{Q_0 + \alpha C_m + \alpha C_r - \alpha S}{2\alpha} \quad (17)$$

Substituting Formula (17) into Formula (5) to obtain the amount of tap water required in the case of cooperation:

$$Q = \frac{1}{2}(Q_0 - \alpha C_m - \alpha C_r + \alpha S) \quad (18)$$

Substitute Formula (17) into Formula (6), and the profit of the water diversion company in the case of cooperation is as follows:

$$\Pi_m^{cooperation} = \frac{1}{2}(Q_0 - \alpha C_m - \alpha C_r + \alpha S)(P_m - C_m + bS) \quad (19)$$

Substituting Formulas (17) and (18) into Formula (8), the overall optimal profit of the two parties is as follows:

$$\Pi^{cooperation} = \frac{(Q_0 + \alpha S - \alpha C_m - \alpha C_r)^2}{4\alpha} \quad (20)$$

From Equation (20) minus Equation (16):  $\Pi^{cooperation} - \Pi^{non cooperation} = \frac{(Q_0 + \alpha S - \alpha C_m - \alpha C_r)^2}{16\alpha} > 0$ . By subtracting Formula (13) from Equation (18), the following can be obtained:  $Q^{cooperation} - Q^{non cooperation} = \frac{1}{4}(Q_0 + \alpha S - \alpha C_m - \alpha C_r) > 0$ . From Equation (17) minus Equation (12) :  $P^{cooperation} - P^{non cooperation} = -\frac{1}{4}(Q_0 + \alpha S - \alpha C_m - \alpha C_r) < 0$ .

Therefore, the overall profit and supply of the cooperative water supply company and the water diversion company are greater than those of the non-cooperative water supply company, and the price of the cooperative water supply plant is lower than that of the non-cooperative water supply company.

Formula (17) derives S and obtains  $\frac{\partial P^{cooperation}}{\partial S} = -\frac{1}{2} < 0$ , that is, the price of water plants decreases with the increase of government subsidy and the rate of reduction is faster than that under non-cooperation. Formula (18) derives S and obtains  $\frac{\partial Q^{cooperation}}{\partial S} = \frac{\alpha}{2} > 0$ , that is, the demand for tap water increases with the increase of government subsidies and the rate of increase is faster than that under non-cooperation.

### (3) Profit distribution of cooperative game based on the Shapley value method

When water companies and water plants cooperate, the distribution of the overall profit benefit is very important. This paper uses the Shapley value method to distribute the overall profit of the supply chain. The expression of the Shapley value method described earlier is as follows:

$$\varphi(v) = \sum_{X \subseteq N} \frac{(|X|-1)!(n-|X|)!}{n!} [v(X) - v(X - \{i\})], \forall i \in N$$

For the supply chain of water diversion project,  $n=2$ , the subset of water diversion

company includes  $X_m = \{\text{water diversion company}\}$ ,  $X_{mr} = \{\text{water diversion company, water supply company}\}$ , and the subset of water supply company includes  $X_r = \{\text{water supply company}\}$  and  $X_{mr} = \{\text{water diversion company, water supply company}\}$ . In the case of  $|X_m|=1, |X_r|=1, |X_{mr}|=2$ , the profit allocation values of the water diversion company and the water supply plant are as follows:

$$\varphi_m(v) = \frac{(1-1)!(2-1)!}{2!} [v(X_m) - 0] + \frac{(2-1)!(2-2)!}{2!} [v(X_{mr}) - v(X_r)] \quad (21)$$

$$\varphi_r(v) = \frac{(1-1)!(2-1)!}{2!} [v(S_r) - 0] + \frac{(2-1)!(2-2)!}{2!} [v(X_{mr}) - v(X_r)] \quad (22)$$

At this time,  $v(X_m) = \Pi_m^{non cooperation}$ ,  $v(X_{mr}) = \Pi^{cooperation}$ , and  $v(X_r) = \Pi_r^{non cooperation}$ . The profit distribution values of the water transfer company and the tap water plant are obtained by substituting Formulas (14), (15), and (20) for Formulas (21) and (22).

$$\Pi_m^{cooperation} = \phi_m(v) = \frac{1}{2}\Pi'_m + \frac{1}{2}[\Pi^{cooperation} - \Pi'_r] = \frac{5(Q_0 + \alpha S - \alpha C_m - \alpha C_r)^2}{32\alpha} \quad (23)$$

$$\Pi_r^{cooperation} = \phi_r(v) = \frac{1}{2}\Pi'_r + \frac{1}{2}(\Pi^{cooperation} - \Pi'_m) = \frac{3(Q_0 + \alpha S - \alpha C_m - \alpha C_r)^2}{32\alpha} \quad (24)$$

The water transfer company and the water supply company cooperate to determine the outlet price of tap water so that the profits of both parties are equal to the profits allocated during the cooperation. Therefore, Formulas (23) and (19) are equal and the water price of the water transfer company under the cooperation situation is as follows:

$$P_m^{cooperation} = \frac{5Q_0 - 5\alpha S + 11\alpha C_m + 5\alpha C_r}{16\alpha} \quad (25)$$

#### 4. Numerical Simulation and Case Analysis

In this paper, data on the water price system of the water diversion project of Tao River are taken as an example to carry out an example analysis (Luo, 2016) [20].

According to the literature (Luo, 2016) [20], the incomplete cost of water distribution per unit of water transfer companies and water plants (specifically the farmers' water users' association) and the bearing capacity and demand of agricultural water are shown in Table 3.

**Table 3.** Cost and demand datasheet of the Tao River Diversion Project (RMB / m<sup>3</sup>).

Parameter	Unit	Incomplete Cost of Yintao Company	Unit Water Distribution Cost of Farmer Water User Association	Maximum Demand
	$C_m$ (RMB/M <sup>3</sup> )	$C_r$ (RMB/M <sup>3</sup> )	$Q_0$ (10000m <sup>3</sup> )	
value	0.28	0.02		4976

Data source: the cost is based on the data from Luo Jinren's study on the water price system of multi-objective water diversion project from the perspective of the supply chain, and the maximum demand is based on the demand forecast data in 2019 from the feasibility report of Taohe River Diversion Project.

Assuming  $\alpha = 1000$ ,  $S$  takes 0, 0.06, and 0.1, respectively. The above data are substituted into the model solution and the calculation results are shown in Tables 4–6.

**Table 4.** Water price profit calculation table 1 of the water diversion project without subsidy.

$S = 0$	$P_m$ (RMB/m <sup>3</sup> )	$P$ (RMB/m <sup>3</sup> )	$\Pi_m$ (10000RMB)	$\Pi_r$ (10000RMB)	$\Pi$ (10000RMB)	$Q$ (10000m <sup>3</sup> )
Non cooperation	2.62	3.81	2733.12	366.56	4099.68	1169
cooperation	0.97	2.64	3416.40	2049.84	5466.24	2338

**Table 5.** Water price profit calculation table of water diversion project when the subsidy amount is 0.06RMB/m<sup>3</sup>.

Calculation Items	$S = 0.06$					
	$b = 1$		$b = 0$		$b = 0.5$	
	Non Cooperation	Cooperation	Non Cooperation	Cooperation	Non Cooperation	Cooperation
$P_m$ (RMB/m <sup>3</sup> )	2.59	0.92	2.65	0.98	2.62	0.95
$P$ (RMB/m <sup>3</sup> )	3.79	2.61	3.79	2.61	3.79	2.61
$\Pi_m$ (10000RMB)	2803.7	3504.64	2803.71	3504.64	2803.71	3504.64
$\Pi_r$ (10000RMB)	1401.8	2102.78	1401.86	2102.78	1401.86	2102.78
$\Pi$ (10000RMB)	4205.57	5607.42	4205.57	5607.42	4205.57	5607.42
$Q$ (10000m <sup>3</sup> )	1169.3	2338.6	1169.3	2338.6	1169.3	2338.6

**Table 6.** Water price profit calculation table of the water transfer project when the subsidy amount is 0.1 RMB/m<sup>3</sup>.

Calculation Items	$S = 0.1$					
	$b = 1$		$b = 0$		$b = 0.5$	
	Non Cooperation	Cooperation	Non Cooperation	Cooperation	Non Cooperation	Cooperation
$P_m$ (RMB/m <sup>3</sup> )	2.57	0.90	2.67	1.00	2.62	0.95
$P$ (RMB/m <sup>3</sup> )	3.78	2.59	3.78	2.59	3.78	2.59
$\Pi_m$ (10000RMB)	2851.27	3564.09	2851.27	3564.09	2851.27	3564.09
$\Pi_r$ (10000RMB)	1425.64	2138.45	1425.64	2138.45	1425.64	2138.45
$\Pi$	4276.91	5702.54	4276.91	5702.54	4276.91	5702.54
$Q$ (10000m <sup>3</sup> )	1169.5	2339	1169.5	2339	1169.5	2339

From the perspective of supply quantity, when  $S=0$ , the  $Q$  under non-cooperation and cooperation are 1169 (10,000 m<sup>3</sup>) and 2338 (10,000 m<sup>3</sup>), respectively. When  $S=0.1$ , the  $Q$  under non-cooperation and cooperation are 1169.3 (10,000 m<sup>3</sup>) and 2338.6 (10,000 m<sup>3</sup>), respectively, the  $Q$  under non-cooperation and cooperation are 1169.5 (10,000 m<sup>3</sup>) and 2339 (10,000 m<sup>3</sup>), respectively, which indicates that the greater the  $S$ , the greater the

Q. When  $S$  is under certain conditions, no matter how much  $b$  is taken, Q is always stable and Q under cooperation is always greater than that under non-cooperation.

From the point of view of water price,  $P$  always decreases with the increase of  $S$ , and non-cooperative  $P$  is always greater than cooperative  $P$ , while  $P_m$  exhibits certain fluctuation with the change of  $b$ . Taking  $S=0.06$  as an example, we can see that  $P$  stays at 3.79 and 2.61 under cooperation and non-cooperation, respectively, while  $P_m$  is 2.65 and 0.98 at  $b=0$ , 2.62 and 0.95 at  $b=0.5$  and 2.59 and 0.92 at  $b=1$ . Under a certain condition of  $S$ , the higher the  $b$  is, the lower  $P_m$  is, and under non-cooperation  $P_m$  is always greater than under cooperation.

From the point of view of profit distribution, with the change of  $S$  from 0 to 0.1, the total supply chain profit  $\Pi$ , the profit  $\Pi_m$  of the water transfer company, and the profit  $\Pi_r$  of the water plant are all increasing. When  $S$  under certain conditions,  $b$  takes different values, there will always be  $\Pi^{cooperation} > \Pi^{non\ cooperation}$ ,  $\Pi_m^{cooperation} > \Pi_m^{non\ cooperation}$ ,  $\Pi_r^{cooperation} > \Pi_r^{non\ cooperation}$ .  $\Pi$ ,  $\Pi_m$  and  $\Pi_r$  are stable because when  $b=0$  increases to  $b=1$ ,  $P_m$  is gradually reduced, thereby maintaining the stability of the internal profit distribution.

## 5. Conclusion

Government subsidy and profit distribution in the supply chain have been among the hot topics in recent years. In the case of low water price tolerance in underdeveloped areas, how to introduce government subsidies to improve water supply and allocate profits in the supply chain are the keys to maintaining a stable and sustainable operation of water transfer projects. In this paper, the model of non-cooperation game and cooperation in the secondary supply chain composed of water transfer company and waterworks is constructed and solved. The water supply, pricing, and profits of the two parties are compared after receiving subsidies.

The results show that with the increase in subsidy amount, the optimal water supply quantity of water diversion project and the profit of all parties are increased, the water price of water plant is reduced, and the water price of water transfer company is also affected by the government subsidy strategy, thereby indicating instability. When the subsidy amount is fixed, the water price of the water plant becomes stable, and the water transfer company can adjust its price according to the subsidy proportion through its dominant position. Regardless of whether the government subsidizes the company, the total profit under cooperation is greater than that under non-cooperation. The profit distribution result based on the Shapley value distribution model can also cause all parties' share profits to become larger than that under non-cooperation, which not only satisfies individual rationality but also the overall rationality and achieves Pareto optimality.

The above conclusion is based on a single water transfer company and a single water plant as the research object. However, the actual situation of the water transfer project is

that a water transfer company faces a complex supply chain network formed by many water plants and farmers' water users' associations. The article also has its limitation in some parameter assumptions. Whether the conclusion is consistent with the supply chain subsidy of the water transfer project requires further study.

## References

- [1] Tabitha M. Fisheries subsidies in China, Quantitative and qualitative assessment of policy coherence and effectiveness. *Marine Policy*. 2016; 68:74–82.
- [2] Lihua H, Liyan CH. The incentive effects of different government subsidy policies on green buildings. *Renewable and Sustainable Energy Reviews*. 2021; 135:110123.
- [3] Dana M, Masayoshi N. Worth the wait: The impact of government funding on hospital emergency waiting times. *Health Policy*. 2020; 124:1340-1344.
- [4] Dongmin K, Shasha L, Yanan W. Market competition, property rights and government subsidies. *Economic Res. J.* 2003; 2: 55-67.
- [5] Hu RC, Zeng WQ, Liu L. A study on the effect of provincial and industry differences of government subsidies-ake listed companies in 6 provinces in Central China as an example. *Chinese J. Manage. Sci.* 2014; S1: 255-266.
- [6] Wang HJ, Li QY, Liu F. Government subsidies: emergency or poverty relief: Empirical Evidence from a sample of loss making companies. *Nankai Business Review*. 2015; 18(5): 42-53.
- [7] Dongmin K, Tianshang L. Has the government subsidy improved the company's performance and social responsibility. *Securities Market Herald*. 2014; 6: 26-31.
- [8] Wu CS, Qian SL, Zhang LJ. The impact of government subsidies on the performance of manufacturing enterprises. *Commercial Res.* 2015; 456: 9-16.
- [9] Gen Q, Hu RX. An analysis on the influencing factors of enterprises receiving government subsidies an empirical study based on industrial enterprise database. *J. Audit & Economics*. 2013; 6: 86-95.
- [10] Wu Q, Ren HB, Gao WJ, Ren XJ, Lao CS. Profit allocation analysis among the distributed energy network participants based on Game-theory. *Energy*. 2017; 118: 783-794.
- [11] Maryam B. A game theoretic approach for pricing petroleum and determining investors' production volume with the consideration of government and intermediate producers. *Sustainable Energy Technologies and Assessments*. 2020; 42: 100825.
- [12] Zhu QH, Dou YJ. Game model of green supply chain management based on government subsidy analysis. *J. Manage. Sci.* 2011; 14(6): 86-95.
- [13] Lu HX. Study on the pricing decision of supply chain of household electrical appliances industry with government subsidies. *Modern Business*. 2015; 27: 13-15.
- [14] Li YD, Zhao DZ, Xia LJ. Government subsidy strategy under vertical emission reduction cooperation of low-carbon supply chain. *Operations Research and Management Science*. 2014; 4: 1-11.
- [15] Zhang XL, Wang JJ. Analysis of government subsidies in the supply chain of new energy vehicles based on Shapley value method. *Soft Science*. 2015; 29(9): 54-58.
- [16] Wang HM, Zhang L, Yang W. Pricing model of water resources supply chain in the east line of South to North Water Diversion Project. *J. Hydraul. Eng.* 2008; 39(6): 758-762.
- [17] Chen ZS. Study on supply chain coordination mechanism of South to North Water Diversion from the perspective of economic benefits and social responsibility. *Resources Science*. 2013; 35(6): 1245-1253.
- [18] Wu LZ, Bai T, Huang Q. Tradeoff analysis between economic and ecological benefits of the inter basin transfer project under changing environment and its operation rules. *J. Cleaner Production*. 2020; 248:119294.
- [19] Wang XZ, Wang YR. Study on the reasonable standard and charging system of water price for the central line of South to North Water Transfer Project. *J. Economics Water Res.* 2015; 1:18-22.
- [20] Jinren L. Research on water price system of multi-objective water transfer project-Analysis Based on the perspective of water resources supply chain. *Price:Theory and Practice*. 2016; 5:61-64
- [21] Jinren L. Basic research on the construction of the mechanism of balancing the interests of stakeholders in water diversion projects- take "Tao River Diversion Project" as an example. Lanzhou University, 2010.
- [22] Weiyi Z. Game theory and information economics, Shanghai people's Publishing House, Shanghai, 1996.

# Developing a Decision Support System for Water Resources Dispatching

Yu FENG<sup>a,b,c</sup>, Jijun XU<sup>b,c,1</sup>, Weirong SHENG<sup>d</sup>, Jitian CHEN<sup>d</sup>, Yang HONG<sup>a</sup>

<sup>a</sup>*School of Earth and Space Sciences, Peking University, Beijing 100871, China*

<sup>b</sup>*Changjiang River Scientific Research Institute of Changjiang Water Resources Commission, Wuhan 430010, China*

<sup>c</sup>*Research Center on the Yangtze River Economic Belt Protection and Development Strategy, CWRC, Wuhan 430010, China*

<sup>d</sup>*Hydrology bureau of Jiangxi Province, Nanchang 330025, China*

**Abstract.** Contradiction between water demand and water supply have a huge impact on social and economic development. This paper presents the development of a water resources dispatch decision support system. The system integrates models related to water dispatch such as streamflow forecast model, water allocation model and water dispatch model. Each model runs as an independent service and is registered in the model platform. The model platform interacts with the service layer and data layer through the model adapter. The model adapter is designed for converting the model input data sent by the service layer and the basic data and observation data queried by the data layer into the format required by the model. In case study, we took the Fu River Basin as an example to demonstrate an application of the system. The system realizes the complete process of data collection, streamflow forecast, water demand declaration, water distribution and water dispatch. User can get the recommended operation plan of the reservoir and the corresponding water supply result through the user interface. Process variables can also be viewed through the system, such as streamflow forecast results and water distribution results, etc. The proposed system can provide technical support and assistance for the decision makers, which also provide an effective demonstration for water resources management in other rivers.

**Key words.** Decision support systems, streamflow forecast, water resource dispatch

## 1. Introduction

Water supply is an essential service that any community need to securely provide [1]. With the development of industrialization and urbanization, the shortage of water resources will become more and more severe due to the growing population and worsening river ecological health [2]. The conflict of water allocation among competing municipal, industrial and agricultural interests is more and more prominent [3-4]. In order to realize the efficient use of water resources and alleviate the imbalance between the supply and

---

<sup>1</sup> Corresponding Author: Jinjun Xu, Changjiang River Scientific Research Institute of Changjiang Water Resources Commission, Wuhan 430010, China; E-mail: [xujj07@163.com](mailto:xujj07@163.com)

This work is supported by the National Natural Science Foundation of China (Grant No. 52009005) and the National Public Research Institutes for Basic R and D Operating Expenses Special Project (no. CKSF2019212/SZ). We also acknowledge the entire development team, without whose help this research could not have been achieved.

demand of water resources, it is necessary to carry out water resources dispatching to meet different needs [5].

In recent years, although researches on water resource dispatching have been carried out, most advances in water resource dispatching models are concealed to the academic arena [6-7]. The usage of water resource dispatching model to actual scenarios remains to be explored [8]. Decision support system (DSS) is an effective way to apply these models to actual scenarios [9]. It helps dispatchers to solve water resource dispatch problems and make decisions through an interactive interface [10]. Researchers proposed the design of DSS based on their own research models [11-13], but most of them are more inclined to the model design. Other researches have improved DDS from the aspect of computer technology, such as [14-15].

In this paper, we focus on the integration of the water resource dispatching models and carried out the research on the development of water resources decision support system. Architecture of the DSS is described. And we also discussed how the model is integrated into the system through the model adapter. This DSS integrated regular scheduling and optimized scheduling models considering different dispatch objects such as flood control, irrigation, water supply, power generation and ecology. These DSS are applied to the water resource dispatching of Fu River in Jiangxi Province, China. In case study, we showed the user interface of the DSS. The system provides an effective means for dispatchers to solve water resource dispatch problems and make decisions.

## 2. Materials and Methods

### 2.1. System Structure

The water resource dispatch decision support system mainly serves the water resource management during the dry season. The system is designed based on a microservice architecture. Applications in the system are split into sub-services. Each service runs in a separate process, and the services use restful APIs to cooperate with each other. Models related to water resource dispatch are also run separately as services. Architecture of the system is shown in Figure 1.

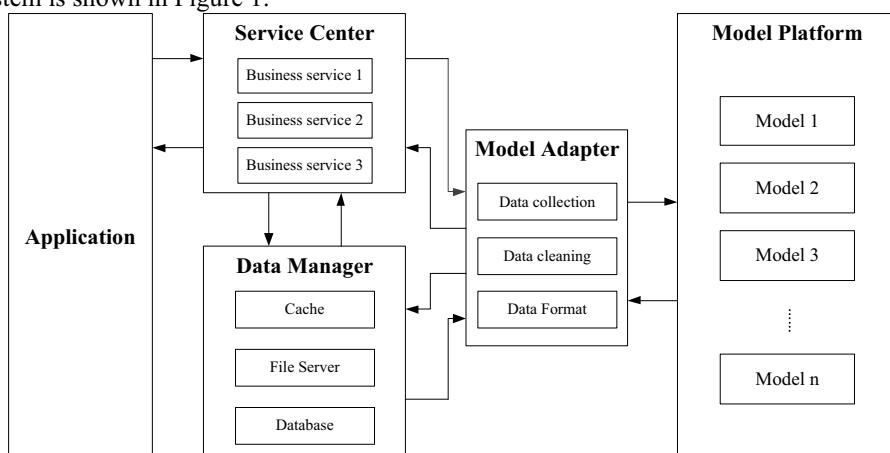


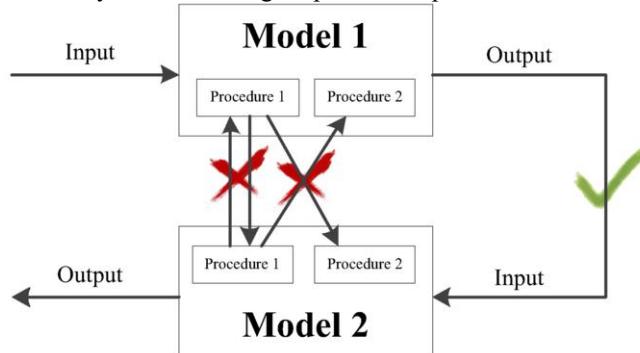
Figure 1. Design of the DSS.

The system is divided into five parts, including Application, Service Center, Data Manager, Model Adapter and Model Platform. The Application provides users with a friendly User Interface, and offers a unified entrance for all users to access the system. The Service Center contains all the business services of the system, such as logical business services, log services, monitoring services, etc. Model calculation requests are also forwarded by the service center. Data Manager is an important component for the system to organize and manage data. Data in the database, Redis and file system is read through the data manager. Model Platform realizes the unified management of the model. Models related to water resource dispatch are registered and managed by the model platform. Model Adapter is the middleware between the model platform and the service layer. The responsibility of model adapter is to collect the necessary basic data and observation data for the model, and convert user input into the format required by the model. The data cleaning component is also implemented in the model adapter.

## 2.2. Model Platform Design

The Model Platform integrates the mathematical models involved in the water resources dispatch process, such as streamflow forecasting model, water allocation model and water dispatch model. The platform also provides model registration service if there are new models that need to be integrated. The platform will provide calling interface for the registered model.

All model integrated in the model platform is designed based on the black box theory. As shown in Figure 2, coupling between models is prohibited. The interaction among different models can only be done through input and output.



**Figure 2.** Interaction among different models.

## 2.3. Data source

Data layer performs unified storage and management of the data in the system. Data from various sources is preprocessed and converted into a standard format, and then stored in the data warehouse. The data stored in the system contains 5 types: real-time observation data, basic information data, business data, spatial data and multimedia data.

- Real-time observation data includes real-time data of monitoring objects such as hydrological stations, reservoirs, rainfall stations, and water users. Types of these data include flow data, water level data, rainfall data, temperature data, water intake data, etc.

- Basic information data contains basic information of model objects, such as the names, addresses, and characteristic values of the reservoirs, hydrological stations, rainfall stations, water users, models, etc.
- Business data is the data generated when the system is running, such as configuration information, user information, operation logs, scheme information, scheme input, model results, etc.
- Spatial data mainly includes basic geographic information data, graphics and remote sensing image data, etc.
- Multimedia data is mainly audio and video data.

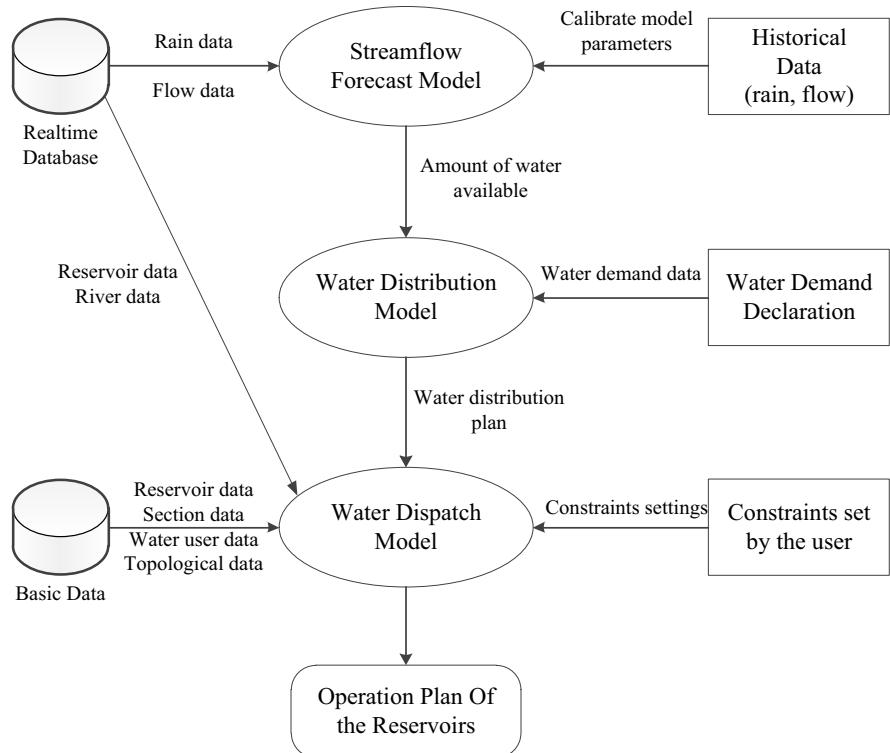
#### 2.4. Models Related to Water Dispatch

The goal of water dispatch is to achieve a balance between water demand and water supply. Future water availability can be estimated through streamflow forecast model [16]. The water demand can be determined by user declaration. Then the water allocation model allocates water consumption according to the weights of different types, including domestic water, industrial water, agricultural water and ecological water. Finally, the operation plan of the reservoirs can be obtained by the water dispatch model. The goal of water dispatch model is to minimize the water supply shortage as much as possible by adjusting the outflow of each reservoir in each time period [17]. The adjustment is mainly in the dry season, while water levels of the reservoirs need to be kept at the flood limit water level during the flood season. Flowchart of water dispatch is shown in Figure 3.

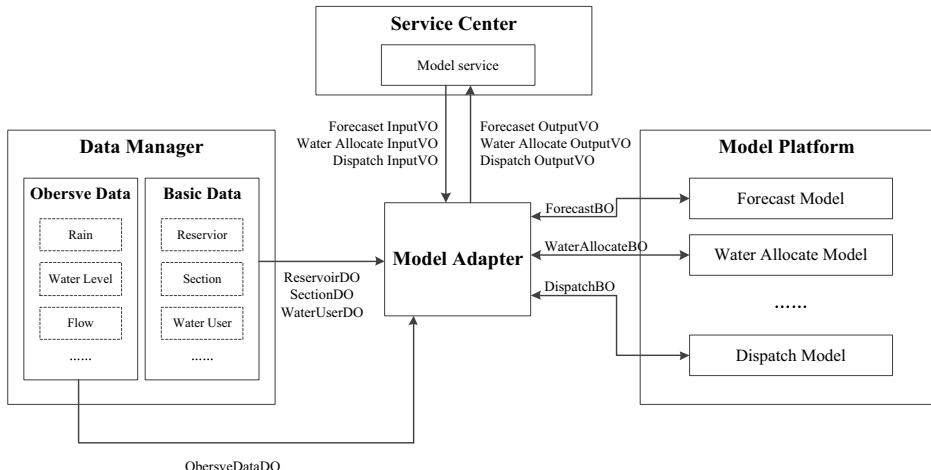
As shown in Figure 3, the parameters of the streamflow forecast model are calibrated based on the historical rain, temperature and flow data. After inputting the observation rainfall and flow data of each station read from the database, the forecast model can predict the streamflow in the future. The water distribution plan is obtained from water demand data and streamflow forecasts data. Then the water dispatch model tries to generate operation plan which meet the water allocation plan as much as possible.

#### 2.5. Model Adapter

When the user uses the water dispatch model, the Service Center sends a model calculation request to the Model Adapter. The model calculation input set by the user will be sent to the Model Adapter as the input view object (VO) along with the request. Then Model Adapter reads the data object (DO) through the Data Manager, including basic information data objects (Basic Data Object) and observation data objects (Observe Data Object). After that, the Model Adapter converts the VO sent by the Service Center and the DO read from the Data Manager into the business objects (BO) required by the models in the Model Platform. Basic Data Object is assembled into Reservoir Object, Section Object, Catchment Object, Water User Object, River Object, etc. according to the different types of model objects. Other object, for example, a Topology Object representing the topological relationship among objects is also assembled. Observe Data Object is transformed into rain series, water level series, flow series, water demand series, etc. These series will be attached to the corresponding model object. After the Model Platform receives the BO, it executes the model calculation. Finally, the model results will be transformed into the VO required by the Service Center through the Model Adapter. Relationships among the system modules are shown in Figure 4.



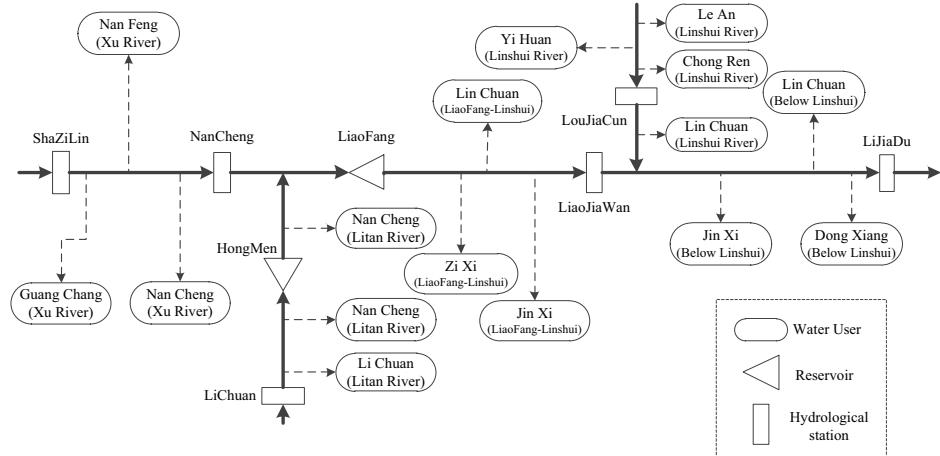
**Figure 3.** Dataflow of water dispatch.



**Figure 4.** Relationship among the system modules.

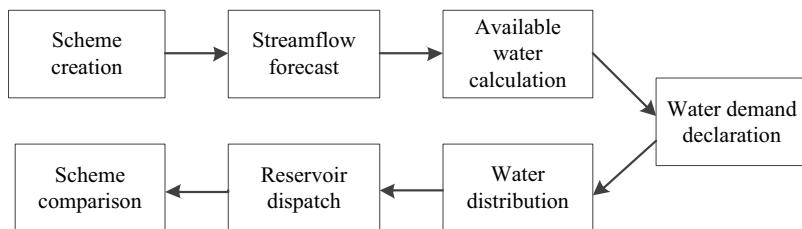
### 3. Results and Discussion

In this section we present an DSS for the water resource dispatching of Fu River. The Fu River is located in the eastern part of Jiangxi Province, China. It is the second largest river in Jiangxi Province. The water resource dispatching involves the water supply of 11 counties. There are two large reservoirs in the basin, with several hydrological stations shown in Figure 5.



**Figure 5.** Topology of the Fu River basin.

In the DSS, every calculation of the model will generate a scheme, which saves the inputs and outputs of the model. The complete water dispatching process starts with streamflow forecasting. Then choose one of the three available water calculation methods provided by the system including the positive algorithm, the reciprocal algorithm and the empirical formula method to calculate the available water. After the water demand is declared and the water allocated is done, the water dispatch model is used to optimize the reservoir operation. If multiple schemes have been made, comparison among these schemes can be done through the system. Flowchart of the water dispatch is shown in Figure 6.



**Figure 6.** Flowchart of the water dispatch.

#### 3.1. Streamflow Forecast Module

In streamflow forecast module, user can set the type of forecast model for each catchment area and the parameters of the corresponding model. After clicking the forecast button, the system automatically collects observation data from the data layer and then do the

prediction. The forecast results are displayed in the form of graphs and tables, such as forecast streamflow, streamflow of the previous year and average annual streamflow. Figure 7 shows one of the user interfaces of the streamflow forecast module. In Figure 7, we can see the monthly streamflow of each section predicted by the forecast model, while some statistical information is also provided.

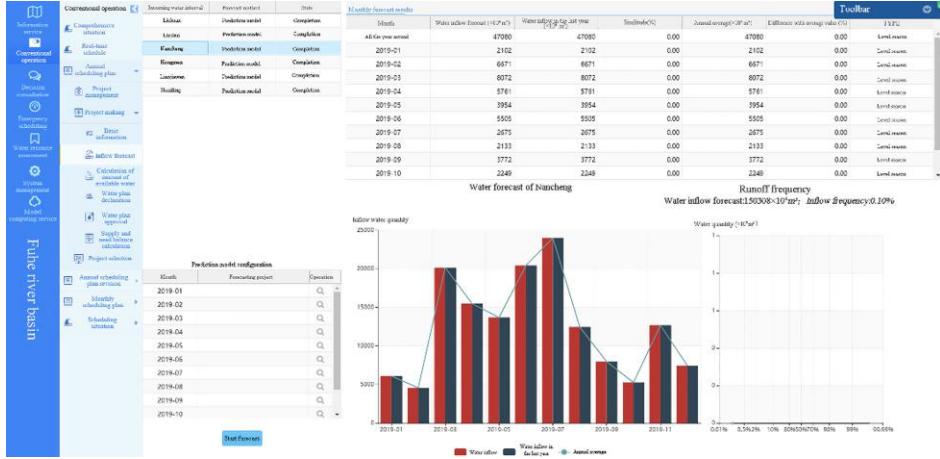


Figure 7. User interface of the streamflow forecast module.

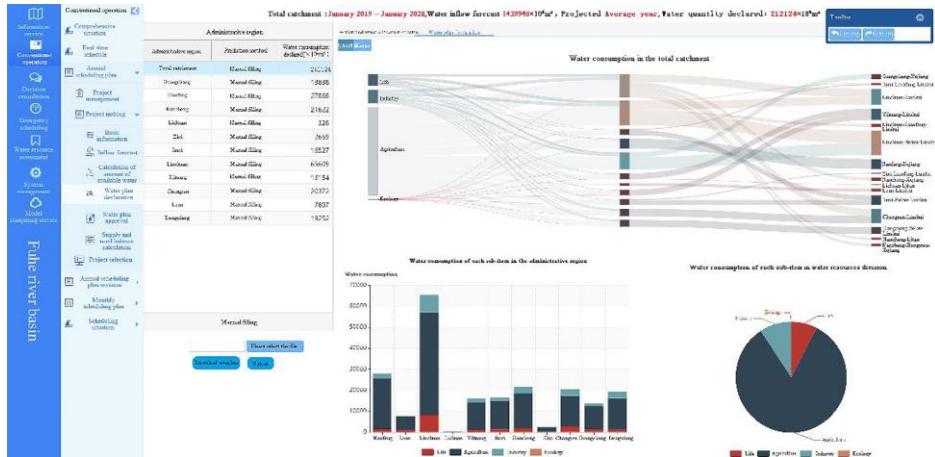


Figure 8. User interface of the water demand declaration module.

### 3.2. Water Demand Declaration Module

In water demand declaration module, user needs to enter the water demand of each water user unit in each time period. User downloads the excel template for water demand declaration, then fill in the corresponding information in the template, and finally upload

it to the system. The system will make statistics for each water user unit and water type. As shown in Figure 8, the system displays the statistical information of domestic, industrial, agricultural and ecological water requirements for each county. Statistical information for the entire basin is also given.

### 3.3. Water distribution module

In water distribution module, system will allocate the available water according to the priority of water user and water type if the available water cannot meet the water demand. Figure 9 shows one of the user interfaces of the water distribution module. The system lists the distribution results of domestic water and ecological water for each water use unit in each month. In general, domestic water demand and ecological water demand have higher priority than that of industrial water demand and agricultural water demand.

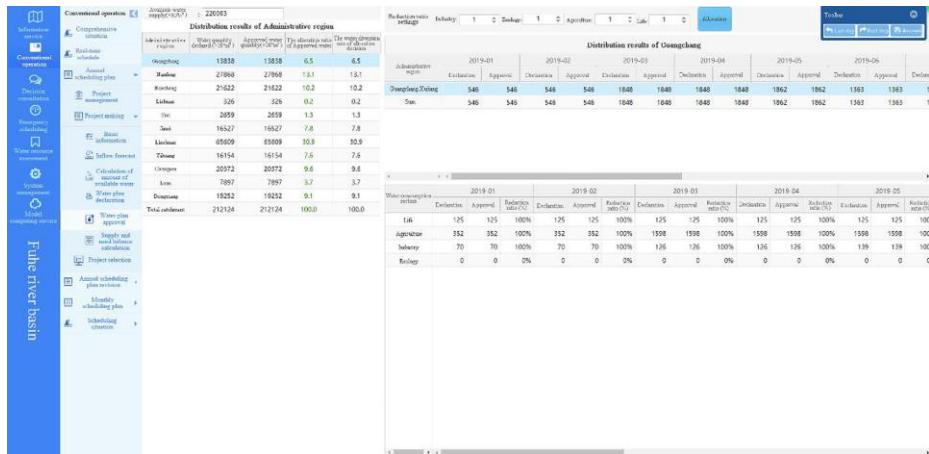


Figure 9. User interface of the water distribution module.



Figure 10. User interface of the water dispatch module.

### 3.4. Water dispatch module

In water dispatch module, user can set the control section constraints and reservoir constraints such as maximum water level, minimum flow, etc. After clicking the calculate button, system will run the water dispatch model to generate a reservoir scheduling scheme. Including the discharge process of the reservoir at each time period, as well as the water supply of each water unit. Figure 10 shows one of the user interfaces of the water dispatch module. In figure 10, the operation process of Hongmeng Reservoir is given, such as water level process, outflow process, etc.

## 4. Conclusion

This paper discussed the design of a water resources dispatch decision support system. The DSS is designed as five parts, including Application, Service Center, Data Manager, Model Adapter and Model Platform. Model Adapter and Model Platform are the featured designs of this article. Model platform integrates mathematical models and provides computing services. The interaction between the model and the system is realized through the Model Adapter. The Application provides users with a friendly User Interface, and offers a unified entrance for all users to access the system. The Service Center contains all the business services of the system, such as logical business services, log services, monitoring services, etc. Data Manager performs unified storage and management of the data in the system. Data from various sources is preprocessed and converted into a standard format, and then stored in the data layer. The system realizes the complete process design from data to water dispatch model calculation.

In case study, we implemented the Fu River Basin Water Dispatching System according to these designs. Using the hydrological and meteorological observation data collected by the system, it is possible to predict the future streamflow of the river. After declaring the water demand of each county and city, the system can apply the water allocation model to allocate water resources to each water user reasonably. Then the operation plan of the reservoir can be obtained by the water dispatch model. User can get the recommended operation plan of the reservoir and the corresponding water supply result. The proposed system can provide technical support and assistance for the decision makers, which also provide an effective demonstration for water resources management in other rivers.

## References

- [1] Fg A , Ac A , Eg B , et al. Dynamic, multi-objective optimal design and operation of water-energy systems for small, off-grid islands - ScienceDirect. Applied Energy. 2019; 250:605-616.
- [2] Jin W , Chang J , Wang Y , et al. Long-term water-sediment multi-objectives regulation of cascade reservoirs: A case study in the Upper Yellow River, China. Journal of Hydrology. 2019; 577:123978.
- [3] Abdulbaki D , Al-Hindi M , Yassine A , et al. An optimization model for the allocation of water resources. Journal of Cleaner Production. 2017; 164(oct.15):994-1006.
- [4] Ahmad I , Fan Z , Liu J , et al. A linear bi-level multi-objective program for optimal allocation of water resources. Plos One. 2018; 13(2):1-25.
- [5] Nematian J , Movahhed S R . An extended multi-objective mixed integer programming for water resources management through possibility theory. Ecological Informatics. 2019; 54:100992.
- [6] Blair, P., Buytaert, W. Socio-hydrological modelling: a review asking "why, what and how?". Hydrology and Earth System Sciences. 2016; 20(1):443-478.

- [7] Li T,Dong Y, Liu Z. A review of social-ecological system resilience: Mechanism, assessment and management. *Science of The Total Environment*. 2020; 723: 138113.
- [8] Pb A , Gg A , Ss B . An actionable hydroeconomic Decision Support System for the assessment of water reallocations in irrigated agriculture. A study of minimum environmental flows in the Douro River Basin, Spain[J]. *Journal of Environmental Management*. 2021; 298:113432.
- [9] Ateş K, Şahin C, Kuvvetli Y, et al. Sustainable production in cement via artificial intelligence based decision support system: Case study. *Case Studies in Construction Materials*. 2021; 15: e00628.
- [10] Quiones-Grueiro M, Prieto-Moreno A, Verde C, et al. Decision Support System for Cyber Attack Diagnosis in Smart Water Networks. *IFAC-PapersOnLine*. 2019; 51(34):329-334.
- [11] Ahmadi A , R Kerachian, Skardi M , et al. A Stakeholder-based Decision Support System to Manage Water Resources. *Journal of Hydrology*. 2020; 589(1):125138.
- [12] Anzaldi G , Rubion E , Corchero A, et al. Towards an Enhanced Knowledge-based Decision Support System (DSS) for Integrated Water Resource Management (IWRM). *Procedia Engineering*. 2014; 89:1097-1104.
- [13] Pierleoni A , Camici S , Brocca L , et al. Climate Change and Decision Support Systems for Water Resource Management. *Procedia Engineering*. 2014; 70:1324-1333.
- [14] Li Y . Towards fast prototyping of cloud-based environmental decision support systems for environmental scientists using R Shiny and Docker. *Environmental Modelling & Software*. 2020; 132: 104797.
- [15] Lazoglu M , Angelides D . Development of a spatial decision support system for land-use suitability assessment: The case of complex tourism accommodation in Greece. *Research in Globalization*. 2020; 2:100022.
- [16] Liu Y , Sang Y , Li X , et al. Long-Term Streamflow Forecasting Based on Relevance Vector Machine Model. *Water*. 2016; 9(1):9.
- [17] Sun J , Lei X , Liang J , et al. A uniform spatial allocation strategy in solving water resources optimization dispatch problem. *MATEC Web of Conferences*. 2018; 246.

# Research on Key Issues of China's Project Water Price

Yanchen ZHOU<sup>a,b</sup>, Xiaoxia TONG<sup>a,c,1</sup>, Zhengjie YIN<sup>a,b</sup>, Lisi XU<sup>a,b</sup>

<sup>a</sup>*Yangtze River Scientific Research Institute of Changjiang Water Resources Commission,  
430010, Wuhan, China*

<sup>b</sup>*Key Laboratory of Water Resources and Eco-Environmental Science of Hubei Province,  
430010, Wuhan, China*

<sup>c</sup>*Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan  
University, Wuhan 430072, China*

**Abstract.** Focusing on China's current water price issues, such as "difficulty in pricing, adjustment, charging, and effectiveness" and the others, the current water conservancy project water supply system has been sorted out in terms of water price formation mechanism, water price composition, influencing factors, price system structure, and pricing technology. Quantitative analysis has been carried out on the problems with prices. At the same time, based on the experience and lessons of similar policies and system reforms at home and abroad, the study puts forward the long-term mechanism of water fee collection and subsidy and the implementation of safeguard measures and recommendations for the water supply price system of water conservancy projects.

**Keywords.** Project water price, pricing mechanism, experience for reference, water price elasticity, affordability

## 1. Research Background

The shortage of water resources in our country and the inherent deficiencies of uneven temporal and spatial distribution, coupled with equal emphasis on waste and pollution, have led to the formation of various water shortages such as management, engineering, and water quality. In recent years, the impact of extreme climate change has increased, which has increased the risk and uncertainty of my country's water resources. Together with food and oil, water resources have become an important strategic resource for sustainable economic and social development. The contradiction between water supply and demand has become a major bottleneck restricting my country's sustainable development. In 2014, the central government put forward the water control policy of "two-handed effort" by the government and the market to ensure water security. The use of market means to adjust the contradiction between water supply and demand, promote the conservation and optimal allocation of water resources, has become a new era of water resources in my country. The key tasks of scientific and efficient management. However,

<sup>1</sup> Corresponding Author: Xiaoxia Tong, Yangtze River Scientific Research Institute of Changjiang Water Resources Commission, 430010, Wuhan, China; E-mail: [ckyttx@163.com](mailto:ckyttx@163.com)

This work is funded by the National Key R&D Program (2017YFC0404504), the National Natural Science Foundation of China (No. 51809009), Key Research Project of Ministry of Water Resources (CKSK2020327/SZ); Special thanks are given to the anonymous reviewers and editors for their constructive comments.

water prices in my country have been low for a long time and cannot effectively reflect changes in water supply costs such as supply-demand relations and price changes[1]. The price elasticity of water supply and demand is lacking, and the role of the market in regulating water supply and demand has not been effectively brought into play. In the past decade, many water supply units and enterprises have generally fallen into a dilemma of "low price and low quality"[2]. Affected by government pricing, public welfare water supply awareness, unscientific cost accounting, water resource fees and sewage treatment fees adjustments, etc., price adjustments are difficult to coexist with improper operation. The responsibilities of the government and enterprises are unclear, and the sharing mechanism is unscientific. The problems are prominent and policy-oriented. Losses and operating losses occur from time to time.

Water supply behavior has a public welfare attribute. The government sets prices low in order to protect the people's livelihood. At the same time, water price adjustment cycles are long and restrictive. In terms of water supply companies, there are problems such as poor management and operation, long-term losses, and insufficient maintenance of the water supply pipe network. The safety of water supply is threatened, and the long-term interests of both the supply and demand sides have been threatened for a long time. At present, China's water supply price does not fully play the role of the two-way market regulation of supply and demand, and there is an imbalance and asymmetry in water supply costs and prices. The boundary between government and market responsibilities is not clear, and the management of urban water supply costs and prices fails, and it is difficult to give full play to the corresponding incentives and restraints.

In order to do a good job in the research on the major issues of price work in the new era and the new normal, take international experience as reference[3-7], clarify the current water supply costs in the water supply industry of water conservancy projects and the implementation of water supply prices and other management problems and causes, seek further measures and methods to solve the problems, and clarify water prices. The direction and goal of the reform have become an important way and inevitable situation to realize the rational and efficient allocation of water resources and alleviate the contradiction between water supply and demand.

## 2. Project Water Price

### 2.1. Water Price Formation Mechanism

With the goal of ensuring water fairness, promoting the rational allocation of water resources and sustainable development, and taking into full consideration the basis of water price determination and comprehensive influencing factors, government departments, water supply and sewage treatment related departments, and water users follow certain guidelines and procedures to This kind of collaborative relationship jointly sets and regulates the work content and process of water supply prices, including cost estimation, organizational system, supervision and inspection, review and approval, implementation and implementation, and related system and policy arrangements. Pricing bodies: water resources prices are determined by the government, industrial and commercial water markets are determined, urban residents' domestic water government markets are jointly determined, and agricultural production and rural water supply governments are determined[8].

The scientific water price formation mechanism must reflect the multiple values of water commodities, the shortage of water resources, the supply and demand of water, and the influence of various factors such as economic and social factors, implement scientific water price pricing methods, and form a scientific relationship between price comparison and price difference. Different water price structures are formed according to different water quantity, water quality, purpose, region, season, peak and valley, and even different water supply projects. A reasonable water price has two meanings: First, its cost should be composed of resource costs, engineering (storage and transportation) costs, water production costs, environmental costs, plus reasonable profits and taxes to form a complete cost that is suitable for sustainable development. Water price; Second, the supply and demand of water resources determined by water price cannot exceed the carrying capacity of water resources and the carrying capacity of water environment. The level of revenue and expenditure determined by water price must ensure the continuous operation of water supply projects and the ability of water users to pay. In other words, the sustainable development of water prices must be able to ensure the sustainable development and utilization of water resources, so that the environment and foundation of water resources regeneration and recycling are not damaged, while taking into account the affordability of water supply projects and the affordability of water users.

## 2.2. Water Price Composition

Water resources have the multiple functions and value characteristics of general natural resources. They are scarce resources. The development and utilization of water resources for economic and social water supply activities not only involve ecological environmental protection and the sustainable use requirements of intergenerational inheritance, but also related to the current economic and social development. The economic benefits, utilization efficiency and consumer rights of the country also produce negative external effects after water and drainage. Therefore, the price of water supply for water conservancy projects should reflect all aspects involved in the development and utilization of water resources. From the analysis of the industrial chain of water supply price formation of water conservancy projects, the water supply of water conservancy projects includes three links: production, operation, and consumption. The formation of terminal prices is the transmission and superposition of costs, prices, and values in each link.

From the perspective of sustainable development, as well as drawing on the successful international experience in the development and utilization of natural resources and the regulation of natural monopoly industries, the price of water supply for water conservancy projects generally adopts full-cost pricing[9,10]. The full-cost water supply price consists of resource costs, engineering costs, environmental costs, taxes and profits. Among them, (1) The cost of resources reflects the ownership of the country, reflects the state as a representative of water resources owned by the whole people to safeguard the rights and interests of common property and social welfare, and reflects the scarcity and water quality differential taxation. This is the economic form of the transfer of all the country's water resources; (2) Project cost is the input cost for turning resource water into product water and making it into the market into commercial water, including engineering fees (reconnaissance, design and construction, etc.), service fees (including operation, operation, management, Maintenance and repairs, etc.) and capital costs (interest and depreciation, etc.). This part of the cost is the most direct cost that constitutes the price of water supply, that is, it is composed of three parts commonly known as raw water cost,

water production cost and water supply operation cost; (3) Environmental cost refers to the manifestation of external costs caused by water resources development and utilization activities, that is, the economic compensation price that causes the reduction of ecological environment functions; (4) Taxes refer to water supply operators that should be paid in accordance with the national tax law and can be included in the water price. Tax; (5) Profit refers to the reasonable income obtained by the water supply operator from engaging in normal water supply production and operation, which is determined by the profit rate of net assets.

Full-cost water supply price = water resource tax + raw water fee (water production cost) + water supply operation cost + sewage treatment fee + business tax + profit

### 2.3. Water Price Pricing Method

Principles: fairness, reasonable distribution and compensation of benefits, dynamic adjustment, promoting the efficient allocation of water resources, reasonable sharing.

The water supply of water conservancy projects has a strong public welfare and a certain degree of profitability. For the business affairs of water production and water supply invested by non-specific assets, the government should gradually transfer specific service functions such as water production and water supply to the main body of the enterprise, and withdraw from urban water supply Specific operations and services in the industry. The cost and benefits of water supply can be shared according to the principle of "who is responsible for funding, who benefits and who invests". The forms of government sharing mainly include infrastructure investment, financial subsidies, and tax incentives. Financial subsidies are one of the main ways. When determining a reasonable amount of financial subsidies, it is necessary to pass cost supervision and review, determine the average production cost of the water supply industry, and define the nature of the company's loss: policy loss or operating loss. Generally, only policy losses are subsidized, and operating losses are not subsidized. Policy-related loss subsidies can consider a combination of direct fiscal subsidies and cross subsidies (implementing tiered water prices to realize the use of riches to make up for the poor).

## 3. Current Water Price

### 3.1. Water Fee Collection

Regarding water fees for water conservancy projects, the national water fee collection rate for large, medium and small water conservancy projects (including some projects directly under the central government and projects in various provinces and cities under separate plans) is about 72%, of which the lowest is 17% in Beijing, and Jiangxi, Sichuan, Hunan, Dalian and directly under the central government. For the project (part of the sample), the water fee collection rate is less than 50%. Guangdong, Jiangsu, Ningxia, Shaanxi, Xiamen, Liaoning, Ningbo, Qingdao and other provinces (autonomous regions, cities) have higher water fee collection rates than 90%. Water charges in various provinces are given in Table 1.

**Table 1.** Collection of water charges in various provinces.

Province	Irrigated area			Reservoir			Rural drinking water	
	Scale	Number	Rate	Scale	Number	Rate	Number	Rate
Central	\	\	\	Large	4			
	\	\	\	Small	0	46%	\	\
	\	\	\	Medium	1			
Beijing	\	\	\	Large	3			
	\	\	\	Small	9	17%	91	72.30%
	\	\	\	Medium	10			
Tianjin	\	\	\	Large	0			
	\	\	\	Small	1	\	\	\
	\	\	\	Medium	7			
Hebei	Large	7		Large	11			
	Small	4	49%	Small	0	75%	2550	92.10%
	Medium	10		Medium	11			
Shanxi	Large	10		Large	7			
	Small	14	83%	Small	28	78%	4027	78.70%
	Medium	132		Medium	61			
Mongolia	Large	5		Large	4			
	Small	10	79%	Small	15	63%	235	88.20%
	Medium	17		Medium	28			
Liaoning	Large	3		Large	2			
	Small	2	49%	Small	0	63%	1017	80.10%
	Medium	12		Medium	15			
Jilin	Large	14		Large	4			
	Small	7	70%	Small	3	69%	3234	89.50%
	Medium	89		Medium	28			
Heilongjiang	Large	14		Large	1			
	Small	2	71%	Small	1	56%	1697	91.80%
	Medium	10		Medium	5			
Jiangsu	Large	6		Large	7			
	Small	0	90%	Small	15	93%	43	90.70%
	Medium	3		Medium	34			
Zhejiang	Large	0		Large	14			
	Small	0	0%	Small	12	76%	1727	98.40%
	Medium	0		Medium	82			
Anhui	Large	13		Large	11			
	Small	33	56%	Small	20	86%	2116	85.20%
	Medium	71		Medium	81			

Province	Irrigated area			Reservoir			Rural drinking water	
	Scale	Number	Rate	Scale	Number	Rate	Number	Rate
Fujian	Large	4		Large	8			
	Small	35	35%	Small	43	68%	2349	93.30%
	Medium	215		Medium	72			
Jiangxi	Large	10		Large	2			
	Small	22	49%	Small	7	28%	1049	93.50%
	Medium	82		Medium	35			
Shandong	Large	13		Large	4			
	Small	0	67%	Small	6	63%	3331	90.00%
	Medium	1		Medium	10			
Henan	Large	35		Large	23			
	Small	28	39%	Small	127	69%	2991	87.60%
	Medium	80		Medium	120			
Hubei	Large	22		Large	23			
	Small	2	67%	Small	8	76%	2755	89.40%
	Medium	36		Medium	59			
Hunan	Large	0		Large	6			
	Small	0	0%	Small	15	34%	3295	30.90%
	Medium	0		Medium	137			
Guangdong	Large	2		Large	8			
	Small	10	32%	Small	17	91%	615	92.60%
	Medium	174		Medium	221			
Guangxi	Large	5		Large	10			
	Small	25	36%	Small	36	75%	105	89.40%
	Medium	79		Medium	37			
Chongqing	Large	0		Large	0			
	Small	1	26%	Small	37	83%	128	90.40%
	Medium	5		Medium	44			
Sichuan	Large	14		Large	0			
	Small	26	35%	Small	44	34%	738	93.20%
	Medium	117		Medium	53			
Yunnan	Large	13		Large	7			
	Small	43	51%	Small	34	70%	4121	75.30%
	Medium	163		Medium	176			
Shanxi	Large	9		Large	2			
	Small	4	90%	Small	6	97%	2761	83.20%
	Medium	38		Medium	9			
Gansu	Large	29	82%	Large	2	82%	2132	89.40%

Province	Irrigated area			Reservoir			Rural drinking water	
	Scale	Number	Rate	Scale	Number	Rate	Number	Rate
Qinghai	Small	31	77%	Small	13			
	Medium	234		Medium	17			
	Large	0		Large	2			
	Small	2		Small	13	69%	4354	65.40%
	Medium	10		Medium	3			
	Large	11		Large	0			
Ningxia	Small	42	86%	Small	33	96%	46	90.10%
	Medium	28		Medium	26			
	Large	53		Large	1			
Xinjiang	Small	11	90%	Small	2	83%	308	90.80%
	Medium	32		Medium	9			

### 3.2. Cost Recovery

Cost recovery of agricultural water supply projects. In large-scale irrigation districts, 33.21% of the national water charges can be guaranteed for operating costs, and projects that can guarantee operating costs plus depreciation account for only 22.4% of the total. Inner Mongolia, Guangxi, Yunnan, Gansu, Xinjiang and other provinces have a relatively high proportion; medium-sized irrigation areas , The proportion of guaranteed operating costs in the country accounts for 26.38%, and the projects that can guarantee operating costs plus depreciation account for only 22% of the total. Provinces and regions such as Qinghai, Heilongjiang, Hubei, and Xinjiang account for more than 50% of the operating costs; small-scale irrigation areas, The proportion of guaranteed operating expenses nationwide is 20.69%, the proportion of projects that guarantee operating expenses plus depreciation is 20.3%, and the proportion of guaranteed operating expenses in Heilongjiang, Hubei, Gansu, Shanxi, Shaanxi and other provinces is more than 50%; in general, the project The smaller, the lower the proportion of projects that guarantee operating costs, and the performance of the Northwest region is slightly better than other regions.

Cost recovery of non-agricultural water supply projects. For large-scale reservoir projects, the proportion of the nation's guaranteed operating costs accounted for 27.40%, and the projects that can guarantee operating costs and depreciation accounted for only 12.33% of the total. Yunnan, Zhejiang, Shaanxi and other provinces have a higher proportion. For medium-sized reservoir projects, 36.86% of the country can guarantee operating costs, while projects that can guarantee operating costs and depreciation account for only 31.57% of the total. Provinces such as Shandong, Zhejiang and Anhui have a higher proportion. For small reservoir projects, 26.59% of the country can guarantee operating costs, and projects that can guarantee operating costs and depreciation account for only 21.97% of the total. The proportion of operating expenses in Fujian, Chongqing and other provinces is more than 50%.

Cost recovery of rural water supply projects. For rural drinking water safety water supply projects, according to the survey of the implementation of tax preferential policies for the construction and operation of rural drinking water safety projects, a statistical analysis of 47,866 rural drinking water safety projects nationwide. Among them, the water

supply price in Sichuan, Beijing, Jiangsu and other regions is greater than the water supply cost, and the water supply price in Ningxia, Xinjiang, Shandong, Inner Mongolia and other regions is less than the water supply cost. Supply cost and water supply price of rural drinking water is shown in Table 2.

**Table 2.** Supply cost and water supply price of rural drinking water.

Province	Water supply cost(Yuan / m <sup>3</sup> )	Water price(Yuan / m <sup>3</sup> )
Beijing	2.99	3.67
Shanxi	2.11	2.24
Mongolia	4.5	2.5
Liaoning	3.54	2.51
Jilin	4.78	3.29
Heilongjiang	1.16	1.51
Jiangsu	1.78	2.27
Zhejiang	5.03	4.16
Anhui	1.79	1.89
Fujian	1.75	1.94
Shandong	3.59	1.97
Henan	2	1.8
Hunan	1.94	2
Guangdong	1.48	1.3
Guangxi	1.38	1.56
Hainan	1.67	1.67
Chongqing	3.01	1.98
Sichuan	2.58	3.32
Guizhou	1.94	2.19
Yunnan	1.94	1.56
Shanxi	2.57	2.41
Gansu	2.83	2.45
Ningxia	3.8	2.3
Xinjiang	3.21	1.66

### 3.3. Problem Analysis

The water price level is low. All personnel of the water pipe unit divide the number of water supply personnel and the number of non-water supply personnel according to a certain ratio; the wages of water supply personnel are not calculated according to the actual wages, but are calculated according to a certain percentage of the average social wages, which are generally smaller than the actual wages; The maintenance cost is calculated according to the actual number of occurrences. The income of the water pipe unit is low, the government funding is insufficient, and the actual maintenance expenditure is much lower than the actual demand.

It is difficult to adjust prices. When local governments adjust prices, they must consider the overall increase in local prices to meet the assessment goals of price control. Due to the characteristics of the public welfare nature of water supply, the government has a heavy psychological burden on price adjustments and the pressure of public opinion. The government would rather not adjust than bear the social risks caused by price adjustments. Water price adjustment does not form a linkage mechanism with related tax and fee adjustments. After any part of the various components of the terminal water price changes (water resource rates, sewage treatment fees, and profits and taxes, etc.), the entire process of cost supervision and review, hearings, and other price adjustments must be followed. The overall water price adjustment is obviously lagging behind other related tax and fee adjustments.

Pricing is difficult. In the survey of water supply costs, labor costs account for a high proportion, are they all included in water supply costs? Single project cost accounting formed a monopoly price, which did not reflect the average social cost. The cost supervision and examination system is not perfect.

It is difficult to collect. At present, China emphasizes poverty alleviation, exempts agricultural taxes and implements various subsidies, the collection of agricultural water prices is inconsistent with agricultural subsidy policies. Water management units, especially water management units that have been transformed into public welfare institutions, are not tied to income and lack the motivation to collect water fees. China's monitoring and measurement facilities are incomplete, and do not match the requirements of charging per party and charging to the household.

#### 4. International Water Price Experience

##### 4.1. "How to Price"

Developed countries generally require cost recovery in the pricing of water supply for water conservancy projects. Irrigation projects led by the federal and state in the United States require all construction and operation and maintenance costs to be recovered and related capital costs (interest) added, while enjoying interest subsidies and preferential treatment. Only part of the construction cost, operation and maintenance cost and capital cost (interest) of rural water supply projects are included in the pricing cost. Grants or special funds are not included in the cost; the United Kingdom implements full cost pricing for all water supply including agricultural water supply. France has subsidized agricultural water use; in India, government subsidies are more in water prices, and agricultural water supply only involves operation and maintenance costs. South Korea has not charged water fees for agricultural water supply in recent years. The above only involves engineering costs, corresponding to domestic resource costs and environmental costs. Pollution costs are generally included in the water price in developed countries, but the calculation of resource costs is not strict. Generally, water environment taxes are used, but it is not common.

##### 4.2. "How Much to Price"

Due to differences in income levels and price levels in typical foreign countries, the overall level of water prices is different. When setting prices, it is generally necessary to consider

the affordability of water users, but water prices, especially agricultural irrigation water prices, are also at a relatively low price, and the elasticity is not significant. Different crop types and irrigation technical conditions also lead to differences in the elasticity of agricultural water use. Among them, the elasticity of artesian irrigation water is relatively high. At present, the highest water prices are in the United States and Germany, with the highest being about US\$8 per cubic meter (including tap water and sewage treatment fees). Mexico's water bill burden rate (water bills as a proportion of household income) is up to 3.7%, typical cities in developed countries are generally within 1%, such as Washington at around 0.7%, and the average household water bill in the UK is stable at 220-300 pounds per year Within the range, only 0.3% of revenue.

#### 4.3. "How to Collect"

In order to promote the collection of water charges, some countries have innovated charging methods. Both the supply and demand parties sign water contracts and pass laws to restrict water users to pay on time. For non-payment and delay in payment, financial sanctions were imposed, and methods such as stopping water supply were even adopted. Most areas in the United States implement a two-part water price, which is collected through a fixed service fee + a metered water fee. There are two charging methods in the UK, one is metered, and the other is not metered, which is based on house size and level. The charges for water users who live in independent villas are higher and the charges for small houses are lower. In addition, certain concessions and exemptions are also given to families of students, families of unemployed workers, bankrupt families, and families with disabilities, including religious sites and charitable organizations. With the improvement of technical means, the implementation of pre-collected water charges began. The difficulty of payment management lies in the collection of agricultural water fees. In order to solve the problem of payment difficulties, countries mainly solve the problem by promoting the autonomy of water supply entities such as irrigation districts and strengthen payment management.

### 5. Suggestion for Water Pricing

At present, in my country's water conservancy and water supply projects, in addition to some of the original funds for project construction, most of the daily operation and maintenance funds of the projects come from water supply income, and the industry implements a self-financing management mechanism. Therefore, recovery of water supply operating costs and a slight profit are the business objectives of the water supply unit. Only by reflecting the cost of water supply and the profit target of the water supply unit, the water price can stimulate the enthusiasm of the water supply unit, improve the efficiency of water supply management, and increase the profitability of water supply.

China's agricultural water price reform and promotion area is still small, even less than 5% in some provinces and cities. It is necessary to tailor measures to local conditions and to formulate agricultural water price reform plans in accordance with the characteristics of local water resources and industrial structure characteristics, promote agricultural water price reform, and expand the area of agricultural water price reform and the area of agricultural water fee charging.

General Secretary Xi Jinping's "3.14" speech deeply discussed the multiple attributes of water and its special status in the national economy and society, and determined that the government and the market must use both the government and the market to determine water prices. Further research clarified the basic concepts of water resource value, water price composition, and water price formation mechanism. In accordance with the instructions of Minister Hubei "Establishing a multi-level water supply price system according to the characteristics of different industries", researched and put forward preliminary ideas for the implementation of regional and water source pricing, and preliminary plans for "industrial profitability, meager living, agricultural irrigation, maintenance, and ecological subsidies". Pricing principles, through a reasonable price tax system and subsidy policies, gradually realize the role of water prices in regulating supply and demand, optimizing the allocation of water resources, and promoting water conservation.

## References

- [1] Li R, Tian DG. Get rid of the dilemma of agricultural water price and respond to the current situation. Reform. 2016; 09:107-114.
- [2] Qin CH. Water Resource Pricing and Its Practice. China Institute of Water Resources and Hydropower Research, Beijing, 2013.
- [3] Guo JJ, Chen S. International experience and enlightenment of water price formation mechanism. China Price. 2021; 3:90-93.
- [4] Mayol A. Social and nonlinear tariffs on drinking water: Cui bono Empirical evidence from a natural experiment in France. Revue d'économie politique. 2018; 6: 1161-1185.
- [5] Arbués F., García-Valinas M. A., Martínez-Espíñeira R. Estimation of Residential Water Demand: a State-of-the-Art Review. Journal of Socio-Economics. 2003; 1: 81-102.
- [6] Grafton RQ, Ward MB, Hang T, Kompas T. Determinants of residential water consumption: Evidence and analysis from a 10-country household survey. Water Resources Research. 2011; 8:1-44.
- [7] Mayol A, Saussier S. Contract renewals, prices and deprivatizations: The case of water in France. Applied Economics. 2021; 21: 2447-2456.
- [8] Michael H, Andreas P. Do political links influence water prices? Determinants of water prices in Germany Utilities Policy.2021; 70: 101184.
- [9] Ding JJ, Wu XB, Yu HP. Research on the Formation Mechanism of Agricultural Water Price Driven by Comprehensive Reform of Water Price-Taking Jingmen, Hubei Province as an Example China Rural Water and Hydropower, 2012; 3:119-122.
- [10] Qin CH, Gan H, Jia L, et al. A model building for water price policy simulation and its application. Journal of Hydraulic Engineering. 2017; 45:109-116.

# The Study of Grading Method and Application Performance of Management Systems of Rural Potable Water Treatment Plants

Xinkai QIU<sup>a</sup>, Cheng LU<sup>a,1</sup>, Yinying ZHOU<sup>b</sup>, Shuyang CHEN<sup>c</sup>

<sup>a</sup>Zhejiang Institute of Hydraulics and Estuary (Zhejiang Institute of Marine Planning and Design), Hangzhou 310020, China

<sup>b</sup>Division of Hangzhou West Lake Aquatic Area Management, Hangzhou 310002, China

<sup>c</sup>Hangzhou Dingchuan Information Technology Co., Ltd, Hangzhou 310016, China

**Abstract.** The application of management systems can solve the manage problems and improve the potable water safety for rural potable water treatment plants. The systems in Zhejiang province, China are studied and the system grading method is proposed as attendance management ( $G_1$ ), basic automatic management ( $G_2$ ), quantity-based automatic management ( $G_3$ ), quality-based intelligent management ( $G_4$ ), quality-based & feedback controlled intelligent management ( $G_5$ ).  $G_3$  to  $G_5$  systems can achieve remote control and  $G_4$ ,  $G_5$  systems can guarantee the finished water quality theoretically. The application performance of the management systems shows  $G_5$  system has the lowest allocated annual cost as 11500 RMB per year when used to service life as 5 years (23.37% of  $G_1$  system). By using  $G_5$  system, the finished water turbidity is below 0.8 NTU, pH is between 7.6 to 8.2, and the qualification rate of residual chlorine is above 92.5%, which performances better than  $G_3$  system with finished water turbidity below 9.7 NTU, pH between 7.3 to 8.2, and the qualification rate of residual chlorine above 88.7%.  $G_5$  system is recommended when the plant is hard to be staffed or the inlet water quality is not good.  $G_3$  system is recommended when the inlet water quality is good or the purchase budget is limited.

**Keywords.** Rural potable water treatment plant, management system, grading method, application performance

## 1. Introduction

The rural potable water treatment plants have to face the manage problems like inadequate equipment maintenance and operators without enough water treatment knowledge[1]. For instance, there were above 9000 villages supplied by rural potable water treatment plants in year 2020 in Zhejiang province, China. It is hard to manage all

<sup>1</sup> Corresponding Author: Cheng Lu, Zhejiang Institute of Hydraulics and Estuary (Zhejiang Institute of Marine Planning and Design), Hangzhou 310020, China; E-mail: lc1001cn@163.com

This work was financially supported by National Key Research and Development Program of China (No. 2019YFC0408801).

the rural plants by human management adequately, and some researches have found rural plants do not have enough managers and one has to take on multiple duties [2].

Using the management systems to replace human managers is an important method to solve the problems and improve the potable water safety. Different management systems have been applied in some plants to monitor the key process of water treatment and make warning [3]. Even more, the computational intelligence has been used in some plants to predict the chemical dosage with good result, instead of the decisions carried out by human experience [4]. There is a necessity to make a comparison study of different management systems, but the previous researches focused more on the performance of a single management system and their grade was not classified. This paper put forward a grading method of the management systems of rural potable water treatment plants. And the application performance of different grades of management systems was analyzed and compared to conduct the application recommendations.

## 2. Materials and Methods

### 2.1. Plants With Graded Management Systems

Different management systems used in rural potable water treatment plants in Zhejiang province were studied to put forward the grading method and to analyze their application performance.

The study of cost performance is based on the plants covered all the grade of management systems. The purchase cost and running cost of the systems were collected and analyzed.

The study of water quality performance is based on three plants with daily supply capacities of 15 tons (*PLANT15*), 100 tons (*PLANT100*) and 600 tons (*PLANT100*). These plants had upgraded the management systems from the grade that cannot guarantee the water quality to the highest grade.

### 2.2. The Analytical Method of Management System Grades

The intelligent systems can judge if the water treatment performance is satisfied and take actions to correct, which is based on monitors, on-line data and lab measures [5], so the equipment and data transmission mode for the systems were analyze in this study. However, the labs were excluded because a great quantity of plants in Zhejiang province do not have it or use water quality monitors for instead.

The equipment can be classified as three types. The first is traditional monitor equipment like camera, alarm and data collection terminal. The second is control equipment like float switch, electromagnetic valve and chemical dosing equipment. The third is detectors and feedback equipment, including programmable logic controller (PLC) and monitors for flow, pressure and water quality.

The data transmission can be classified as two modes. The first use SIM card with 2G or 4G internet to transmit pictures or videos and help monitoring the attendance of managers. The second use cable network to transmit data and command stably, which can achieve remote control.

The combination of equipment and transmission mode can achieve different management functions. and the management system grades can be classified by the functions.

### *2.3. The Analytical Method of Application Performance*

The application performance includes costs and finished water quality of different management systems.

The cost analysis is based on the purchase cost of management equipment and running cost of labor and materials incurred for the systems. The systems studied with different grades have the same designed serviceable life as 5 years, and the purchase and running cost were separated into the service period to calculate the allocated annual cost.

Total coliforms, turbidity and pH have the highest ineligible rate in finished water of rural potable water treatment plants in Zhejiang province [6], which are most concerned and used to represent the water treatment performance. Therefore, the data of residual chlorine, turbidity and pH was collected by water quality monitors and analyzed in this study. The finished water quality is compared with the Standards for Drinking Water Quality of China (GB 5749-2006) to judge if it is qualified [7]. Also, the inlet water with low turbidity (below 10 to 30 NTU) [8] is hard to treat. This study analyzed the inlet water turbidity below or above 15 NTU separately, if the data is monitored in a large range.

## **3. Results and Discussion**

### *3.1. The Grading Method of Management Systems*

Depends on the equipment and data transmission mode, the functions of the management systems can be classified into five grades.

The first and least intelligent grade is attendance management. By equipping camera and transmit data with SIM card, the plant can capture the pictures to record the attendance and alarm the managers.

The second grade is basic automatic management. By equipping automatic control equipment and transmit data with SIM card, the plant can automatic start or stop. For example, when the water in clean water basin is enough, the floating switch will cut off water produce [9].

The third grade is quantity-based automatic management. By equipping automatic control equipment, inlet flow monitor and PLC, the plant can change the chemical dosage to fit the inlet flow [10]. From this grade and higher grades, the plants can be remote controlled with cable network.

The fourth grade is quality-based intelligent management. By adding pressure monitors and finished water quality monitors over the third grade. The plant can increase the chemical dosage to make the finished water turbidity back to be qualified, and backwash intelligently to decrease the pressure difference from inlet and finished water. However, this system cannot retrace the increased chemical dosage settings when the inlet water quality gets better or the finished water gets easier to be qualified.

The fifth and most intelligent grade is quality-based & feedback controlled intelligent management. By adding inlet water quality monitors over the fourth grade,

the plant can adjust the treatment settings by monitoring the water quality, and also retrace it when the inlet water quality has a better feedback.

Depending on the functions, the grading method can be set as five levels given in table 1.  $G_1$  to  $G_3$  systems are unable to monitor or deal with the change of water quality, so the finished water qualities for these three grades are not guaranteed theoretically.

**Table 1.** The grading method of management systems for rural potable water treatment plants.

Grades	Management functions	Management contents	Control methods	Finished water quality
$G_1$	Attendance management	All contents		
$G_2$	Basic automatic management	Chemical replenish, dosage adjustment	Field management	Quality not guaranteed
$G_3$	Quantity-based automatic management	Chemical replenish,		
$G_4$	Quality-based intelligent management	dosage adjustment	Remote management	Quality guaranteed
$G_5$	Quality-based & feedback controlled intelligent management	Chemical replenish		

### 3.2. The Application Cost of Management Systems

The equipment and the cost are given in table 2. The running cost is given in table 3. In the locality, the personnel cost for field manager is 2000 RMB·month<sup>-1</sup> for  $G_1$  system and half for  $G_2$  system. The personnel cost for engineer to adjust the chemical dosage is 2000 RMB·month<sup>-1</sup>·plant<sup>-1</sup>. The system with residual chlorine monitors need 260 RMB·month<sup>-1</sup> for the replacement of detection reagent.

**Table 2.** The purchase cost of management systems.

Grades	Controlling equipment and monitor equipment	Others	Cost (RMB)
$G_1$	Data collection terminal	Camera	30000
$G_2$	Flow switch, float switch, electromagnetic valve, electromagnetic relay, chemical dosing equipment	Video camera	80000
$G_3$	$G_2$ system adding PLC, Inlet flow monitor		120000
$G_4$	$G_3$ system adding pressure monitor, finished water quality monitor (turbidity, pH, residual chlorine)	Video camera, intruder alarm,	170000
$G_5$	$G_4$ system adding inlet water quality monitor (turbidity, pH)	voice broadcast	210000

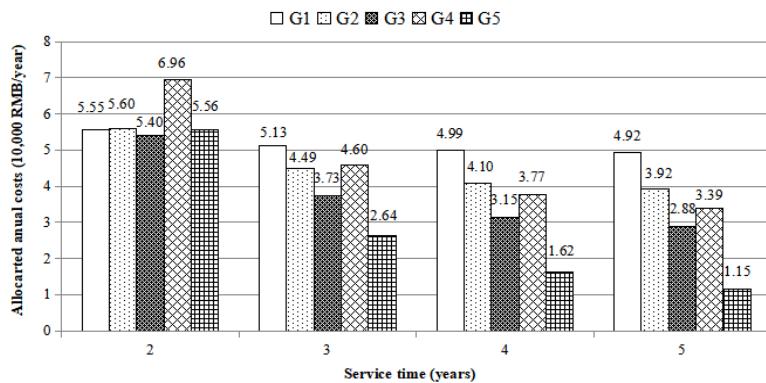
**Table 3.** The running cost of intelligent management systems.

Grades	Manpower demand	Personnel cost (RMB·month <sup>-1</sup> )	Material cost (RMB·month <sup>-1</sup> )	Running cost (RMB·year <sup>-1</sup> )
$G_1$	Field manager & Engineer	4000	/	48000
$G_2$	Field manager & Engineer	3000	/	36000
$G_3$	Engineer	2000	/	24000
$G_4$	Engineer	2000	260	27120
$G_5$	/	/	260	3120

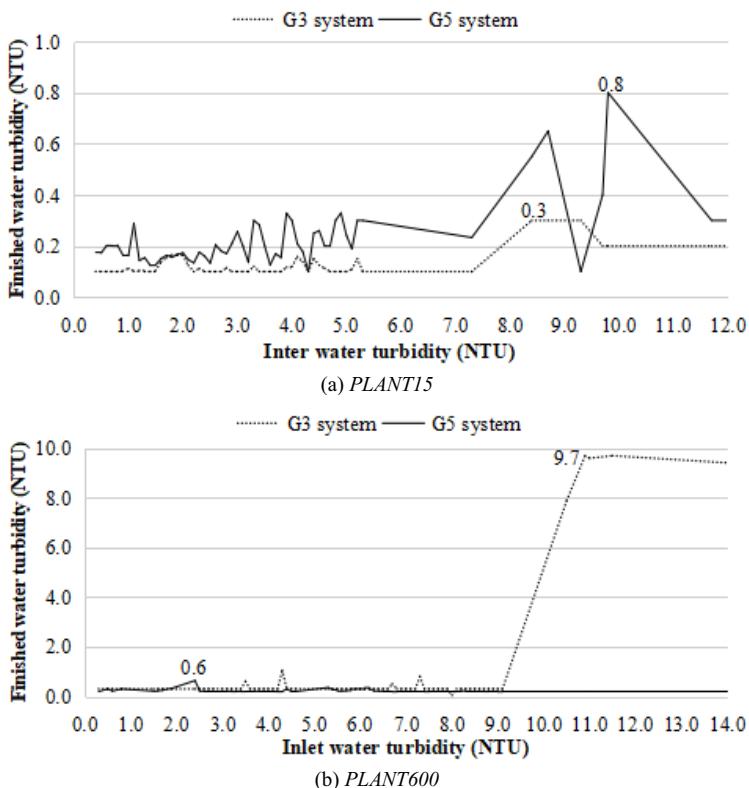
$G_4$  system can guarantee the finished water quality but still needs engineer to adjust chemical dosage settings, and have extra cost for detection reagent, so its running cost is higher than  $G_3$  system. Other grade of systems has lower running cost than the lower grades. After allocating the purchase cost into the service period happened, the allocated annual cost is given in figure1, which shows  $G_3$  system has the lowest cost when used for 2 years, and  $G_5$  system has it when used for 3 and more years.

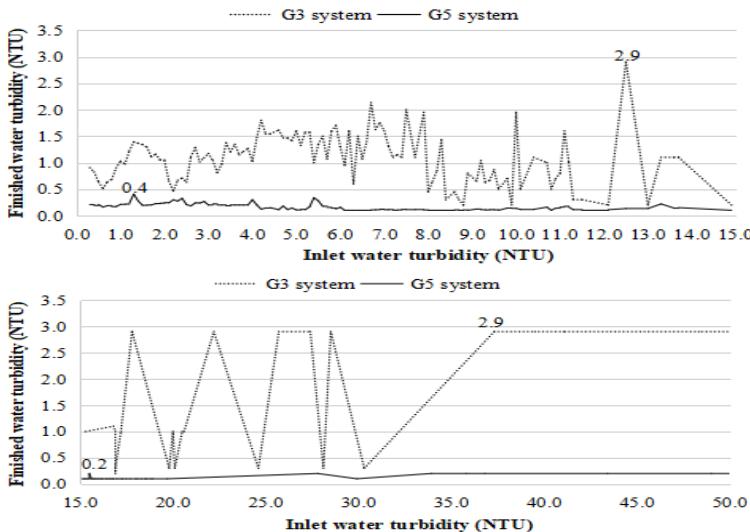
### 3.3. The Finished Water Quality Performance of Management Systems

Because  $G_1$  to  $G_3$  systems cannot guarantee the finished water quality, the water qualities with  $G_3$  and  $G_5$  from three plants were analyzed to show the enhance of water quality by the upgrade of management systems.



**Figure 1.** The allocated annual costs of management systems.



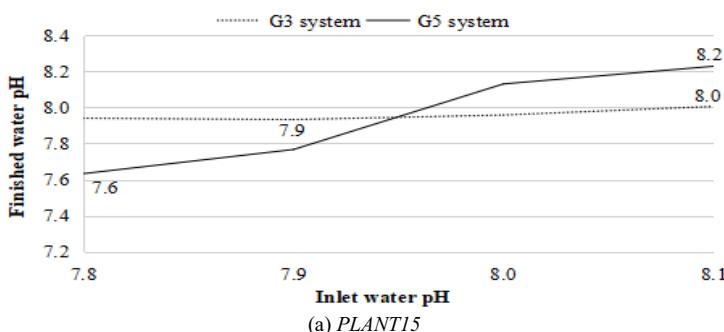
Figure 2. The water turbidity with  $G_3$  &  $G_5$  systems.

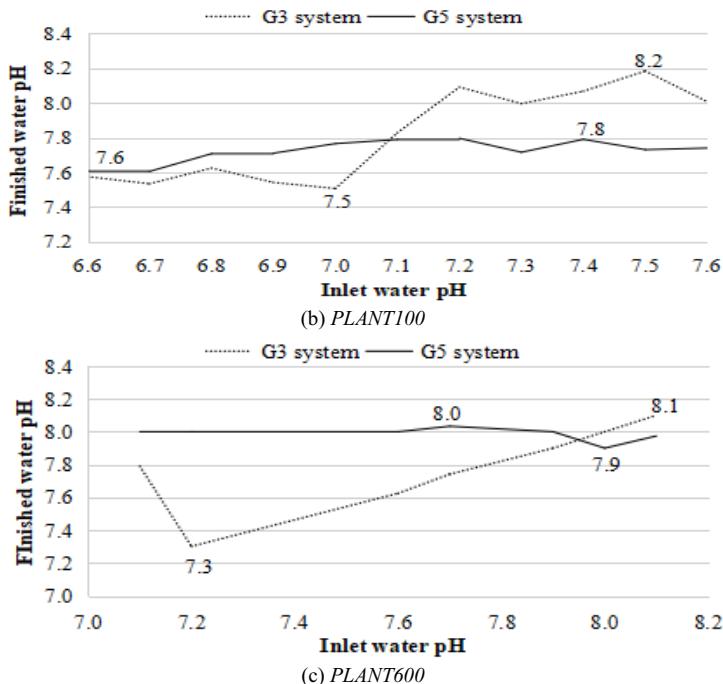
### 3.3.1 Turbidity

The inlet and finished water turbidity for three plants are given in figure 2. It shows *PLANT15* has lower inlet water turbidity and  $G_5$  system does not show any improvement. *PLANT100* and *PLANT600* with  $G_5$  systems have lower finished water turbidity than those with  $G_3$  systems. It should be noticed that when the inlet water turbidity is higher than 9 NTU in *PLANT600*,  $G_3$  system is unable to monitor it or change water treatment settings, so its finished water turbidity can not meet the standard (below 3 NTU).

### 3.3.2 pH

The inlet and finished water pH for three plants are given in figure 3. It shows all the finished water pH with  $G_3$  or  $G_5$  systems can meet the standard (between 6.5 to 8.5). *PLANT15* has a stable inlet water pH, and  $G_5$  system does not show any enhance. *PLANT100* and *PLANT600* with  $G_5$  systems has more stable finished water pH than those with  $G_3$  systems.





**Figure 3.** The treatment performance of pH with  $G_3$  &  $G_5$  management systems.

### 3.3.3. Residual Chlorine

The finished water residual chlorine for three plants are given in table 4. It shows the residual chlorine qualification rates with  $G_5$  systems are higher than those with  $G_3$  systems. Especially for small plants like *PLANT15* with  $G_3$  system, its chemical dosage is based on manual settings and hard to meet the standard.

The average residual chlorine with  $G_3$  system is  $1.1 \text{ mg}\cdot\text{L}^{-1}$ , while it is more stable and approaching the standard value ( $0.3 \text{ mg}\cdot\text{L}^{-1}$ ) with  $G_5$  system. The finished water having lower and qualified residual chlorine with  $G_5$  system is more acceptable for rural water users.

**Table 4.** The treatment performance of residual chlorine with  $G_3$  &  $G_5$  management systems.

Plants	Residual chlorine with $G_3$ system		Residual chlorine with $G_5$ system	
	Average ( $\text{mg}\cdot\text{L}^{-1}$ )	Qualification rate (%)	Average ( $\text{mg}\cdot\text{L}^{-1}$ )	Qualification rate (%)
<i>PLANT15</i>	1.1	88.7	0.3	92.5
<i>PLANT100</i>	0.4	94.3	0.4	97.8
<i>PLANT600</i>	0.7	97.1	0.6	98.9

## 4. Conclusions

The management systems for rural potable water treatment plants can be graded as attendance management ( $G_1$ ), basic automatic management ( $G_2$ ), quantity-based automatic management ( $G_3$ ), quality-based intelligent management ( $G_4$ ), quality-based &

feedback controlled intelligent management ( $G_5$ ).  $G_3$  to  $G_5$  systems can realize remote control,  $G_4$  and  $G_5$  systems can guarantee the finished water quality.

$G_5$  system has the lowest allocated annual cost as 11500 RMB per year (23.37% of  $G_1$  system) when used for 3 and more years.  $G_3$  system has only 57.14% of the purchase cost of  $G_5$  system, and has the lowest allocated annual cost as 54000 RMB per year (77.59% of  $G_4$  system) when used for 2 years.

$G_5$  system can improve the performance of finished water quality. The finished water turbidity is all qualified and below 0.8 NTU with  $G_5$  system, while it is not all qualified and below 9.7 NTU with  $G_3$  system. The finished water pH is between 7.6 to 8.2 with  $G_5$  system., while it has a wider deviation between 7.3 to 8.2 with  $G_3$  system. The qualification rate of residual chlorine is above 92.5% with  $G_5$  system, while it is only above 88.7% with  $G_3$  system.  $G_3$  system may has better finished water quality if the inlet water quality is good and stable, take this study as example, the inlet water turbidity is below 12 NTU and pH is stable from 7.9 to 8.0.

$G_5$  system is recommended when the plant has no manager, the inlet water quality is limited and the finished water quality is highly demanded.  $G_3$  system is recommended when the inlet water quality is good, the purchase budget is limited.

## References

- [1] Momba MNB, Tyafa Z, Makala N. Rural water treatment plants fall to provide potable water to their consumers: the Alice water treatment plant in the Eastern Cape province of South Africa. *J. Sci.* 2004; 100(5-6):307-310.
- [2] Wu XM, Sun QS, Shan J. Research on Post Arrangement and Fixed Number of Persons of Rural Water Supply Enterprise. *China Rural Water and Hydropower*. 2014; 1:61-63.
- [3] Wu J, Hao XW, Yang CJ. Programming and Application of Online Monitoring and Warning System for Rural Drinking Water. *Zhejiang Hydrotechnics*. 2013; 41(3):1-3.
- [4] Fabio CRS, André FHL, Cleber G Dias, et al. Intelligent system for improving dosage control. *Acta Scientiarum: Technology*. 2017; 39(1):33-38.
- [5] Dan S, Gopal A, Cooper HL, et al. Performance management of small water treatment plant operations: a decision support system. *Water and Environment Journal*. 2017; 31(3):330-344.
- [6] Ren BZ, Deng RJ. Safety of Rural Drinking Water and Its Counter measures. *China Safety Science Journal*. 2008; 5:11-17.
- [7] He XJ, Wang B, Qiu C, et al. Water quality investigation and analysis for rural drinking water in Zhejiang Province China. *Water Resources*. 2019; 11:40-42.
- [8] Zhao XD. Treatment of Low Temperature and Low Turbidity Water. *Journal of Shanghai University of Electric Power*. 2009; 25(05):487-490.
- [9] Li JH. Design and application of remote monitoring system for centralized water supply. *Modern Electronics Technique*. 2013; 17: 96-97.
- [10] MO RB. Analysis and Solution for Water Quality Problems in Rural Water Supply Project. *Water Purification Technology*. 2021; 40(1):146-149.

# Application of Analytic Hierarchy Process (AHP) and Simple Additive Weighting (SAW) Methods in Mapping Flood-Prone Areas

Yu CHEN<sup>a,b,1</sup>

<sup>a</sup>*State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Chengdu 610065, China*

<sup>b</sup>*Institute for Risk and Disaster Reduction, University College London, Gower Street, London WC1E 6BT, United Kingdom*

**Abstract.** The first step of formulating flood risk management strategies is to identify the flood at-hazard areas. This study aims to map flood-prone areas with different hazard levels in the Dadu River basin, using simple additive weighting (SAW) and analytic hierarchy process (AHP) methods and the geographic information systems (GIS) tool. The grid map of selected criteria, rainfall, topography, drainage, and the usage of land were processed and applied to estimate the flood hazard index (FHI) values in the basin in the GIS environment. The resultant map illustrates the spatial distribution of basin-scale flood at-hazard areas, can be used as powerful guidance of implementing preventing and alleviating flood risk for decision-makers and managers, and extended application in other basins or disaster fields.

**Keywords.** Flood hazard, flood-prone areas, SAW, AHP, GIS

## 1. Introduction

The flood is a multi-attribute natural hazard and often causes catastrophic damages and enormous impacts to the human, society, economy and environment, and has become a worldwide problem. The increasing flood occurrence necessitates the development of flood management, where flood hazard evaluation is one important part with the research purpose of obtaining the spatial distribution of flood-prone areas, which is a basis of implementing effective management decision-making.

Numerous hazard assessment methods have been developed, which include statistical analysis, uncertainty analysis, and multi-criteria evaluation (MCE), etc. [1]. In recent years, more and more researchers pay attention to the MCE, especially the geographic information system (GIS)-based MCE method, which is a useful tool for

<sup>1</sup> Corresponding Author: Yu Chen, State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Chengdu 610065, China; E-mail: [rainchen393@hotmail.com](mailto:rainchen393@hotmail.com)

This work was supported by Sichuan Science and Technology Program [grant number 2019YFH0140]; International Visiting Program for Excellent Young Scholars of SCU; and the National Natural Science Foundation of China [grant numbers 41501554].

processing spatial data and can incorporate all relevant types of consequences [2]. Jonkman et al. (2008) [3], Meyer et al. (2009) [4], Wang et al. (2011) [5], Papaioannou et al. (2015) [6] successively researched its application in flood risk assessment.

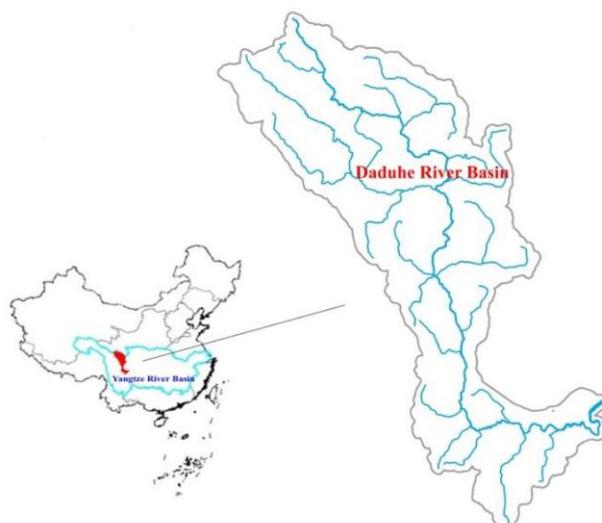
Analytic hierarchy process (AHP) with the major advantage of simple GIS integration is the most commonly used MCE method in the flood risk management field [7]. SAW is also a broadly applied MCE approach, and its combination with GIS can carry out the spatial analysis of the evaluation and decision situations [8]. The researches of [9-11] have successively identified the practicability of the integration of AHP, SAW and GIS. Therefore, this study applies SAW and AHP methods to establish a GIS-based MCE model for mapping flood hazard areas with different classes to support flood risk management decision-making.

## 2. Methodology

### 2.1. Study Area

Dadu River basin is located in Sichuan Province, Southwestern China (Figure 1), covers an area of 77400 km<sup>2</sup>, and has been heavily hydropower-developed in the form of cascade reservoirs. The basin has a complex climate and geological environment, and is a global climate change sensitive area and the main geological activity area.

The basin is prone to flooding due to heavy precipitation. Since 1956, several historical flood surveys have been carried out in the basin. The surveyed historical flood years are 1786, 1892, 1904, 1939, and 1955. Except for the mountain collapse-induced flood in 1786, the biggest flood in the investigation is the 1904 flood, followed by the 1939 flood. For the 1904 historic flood event, caused by rare and long-term rainfall in the upstream of the basin, the flood return periods in the upper reaches, the middle reaches and the lower reaches are more than 200-year, 100-year, and more than several-decade, respectively. The heavy flood in 1939 was mainly formed by the rainstorm in the middle and lower reaches.



**Figure 1.** Location of the Dadu River basin.

## 2.2. Criteria Identification

The criteria selection is a vital step for flood hazard evaluation. In the study, five indicators, maximum five-day precipitation (M5DP,  $C_1$ ), digital elevation model (DEM,  $C_2$ ), slope ( $C_3$ ), land cover use (LCU,  $C_4$ ), and distance to the river (DTR,  $C_5$ ), are selected based on the previous studies [12-16] and the actual situation of the study area.

M5DP is a representative rainfall intensity indicator and abstracted from the meteorological-station data (<http://cdc.cma.gov.cn/>) in the basin. DEM is the most important criterion for flood hazard assessment, can be extracted from the 30×30m ASTER GDEM data (<http://www.gscloud.cn/>), and so does slope. The slope is crucial to regulate the flow of surface water, and partially controls the infiltration process. The regions with lower slope are prone to flood hazard, while higher slope provides an easy passage to pass away the flood. Areas closed to rivers are much more likely to suffer more frequent floods. To quantify this indicator, sequential buffers were hence created along the rivers and corresponding values were assigned, for instance, the rivers are set to 0, and similarly, the regions far more away from rivers are assigned larger values.

The river system is extracted based on ASTER GDEM data. LULC illustrates the land cover, reflects varying degrees of flood interception, and can be downloaded from <http://www.gscloud.cn/>.

To quantify these criteria and delineate different flood hazard zones in the basin, indicator data sets garnered from various sources should be input into GIS and converted to spatial criteria map layers with the same spatial resolution (30m grid cell size).

## 2.3. AHP and SWA Methods

AHP as the most common used MCE method is utilized to calculate criteria weights. Based on a preference matrix, all evaluated criteria are compared against each other in pairs, and assigned weights correspondingly. Then the assessment results can be derived by aggregating the criteria with determined weights in selected combination methods [2].

The pairwise comparison matrix  $A = [a_{ij}]$ , ( $i=1,2,\dots,m$ ,  $j=1,2,\dots,n$ ) is defined as follows:

$$A = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ 1/a_{12} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1/a_{1n} & 1/a_{2n} & \cdots & 1 \end{bmatrix} \quad (1)$$

Where the entry  $a_{ij}$  expresses how much the criteria  $x_i$  is preferred to criteria  $x_j$ , if  $a_{ij} = a$ , then  $a_{ji} = 1/a$ ,  $a > 0$ . If all criteria are already known,  $a_{ij}$  equals to the ration of relative weights of criteria  $x_i$  and  $x_j$ ,  $w_i / w_j$ . To determine the relative weights, the pairwise judgement is made by decision makers, which is based on the rule of Saaty (1977) [17] with a 9-point scale from 1 to 9 in Table 1. The consistency ratio (CR) calculated in terms of Eq. (2) is used to check the consistency of judgement. If CR is less than 0.1, then the judgement can be accepted.

$$CR = CI / RI \quad (2)$$

CI is consistency index and denoted as:

$$CI = (\lambda_{\max} - n)/(n-1) \quad (3)$$

where  $\lambda_{\max}$  is the largest eigenvalue derived from the paired comparison matrix,  $n$  is the number of criteria.

In Eq. (2), RI is random index, and obtained according to Table 2 defined by Saaty (1977) [17].

**Table 1.** Scales for pairwise comparisons.

Intensity of importance	Definition
1	Equal importance
2	Weak
3	Moderate importance
4	Moderate plus
5	Strong importance
6	Strong plus
7	Demonstrated importance
8	Very importance
9	Extreme importance

**Table 2.** Random index (RI).

<i>n</i>	1	2	3	4	5	6	7	8	9
<i>RI</i>	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45

SAW is based on the weighted summation concept to estimate the flood hazard index (FHI) value, which is specified in terms of the following formulas.

$$FHI = \sum_{j=1}^n w_j r_{ij} \quad (4)$$

$$r_{ij} = \frac{x_{ij} - \min x_{ij}}{\max x_{ij} - \min x_{ij}} \quad (5)$$

where  $x_{ij}$  is the value of  $i$ th alternative on  $j$ th criteria,  $r_{ij}$  is the normalized value of  $i$ th alternative on  $j$ th criteria,  $\max x_{ij}$  and  $\min x_{ij}$  are the maximum and minimum values of  $x_{ij}$ , respectively,  $w_j$  is  $j$ th criteria weight calculated by AHP method, and  $n$  is the criteria number.

### 3. Results and Discussion

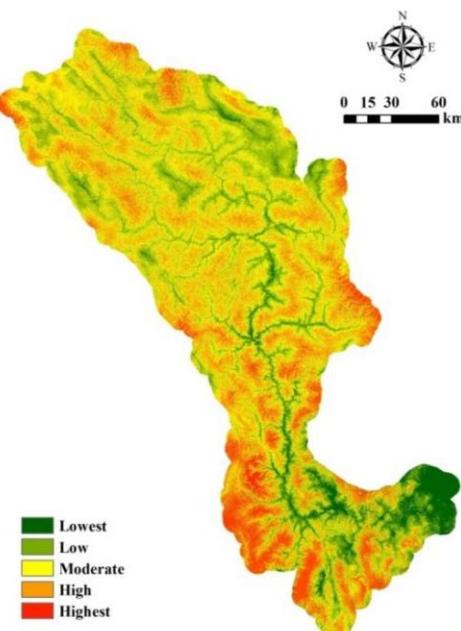
Constructing a pairwise matrix is the necessary step of using the AHP method. Then, criteria weights and the CR can be calculated (Table 3). The resultant CR is less than 0.1, which indicates the comparison is consistent and the assigned weights are appropriate. It can be seen that DEM is assigned the highest weight, followed by slope, DTR, and M5DP. LUC is identified as less important. This matches the fact that the flood is mainly driven by rainfall, and the driving force will increase with the facilitation of topography to runoff and the de-vegetation of landscapes.

After using the SWA method to calculate FHI values based on the weighted criteria layers, the derived flood hazard map was reclassified into five levels of lowest, low, moderate, high, and highest (Figure 2) by using the natural breaking method, and the corresponding areas of each level account for 6%, 24%, 38%, 24%, and 8% of the basin, respectively.

The results demonstrated that the east part of the downstream presents low to lowest flood hazard potentiality as the elevation and slope is very high comparing with other regions. About 38% of the basin area is distributed in the moderate flood hazard zone, which is the overriding hazard level and spread over a wide range, without a uniform distribution pattern. The highest flood potential area is mainly distributed in the midstream of the basin where a rainfall center is located, and elevation and slope are lower, which plays a crucial role to control the flood potentiality of an area. The area zoned as the highest flood hazard level needs great attention of risk administrative organization to prevent more serious flood hazard situations in the future.

**Table 3.** Pairwise comparison matrix, consistency ration and weights of the criteria.

	LUC	M5DP	DTR	slope	DEM	Eigen values
LUC	1					0.0548
M5DP	4	1				0.1206
DTR	4	2	1			0.1580
slope	4	3	3	1		0.2884
DEM	4	3	3	2	1	0.3782
Consistency ration: 0.07						



**Figure 2.** Flood hazard level map in the Dadu River basin.

Assessing the basin-scale flood hazard is an effective way for accumulating correlated disaster risk information to support risk management decision-making of preventing and alleviating the flood disaster. To identify the flood hazard-prone areas, risk researchers and decision-makers have developed several approaches to predict the spatial distribution of flood hazard levels, wherein how to discern and utilize the most optimal method to generate the flood hazard map for specific areas is the common difficulty. In this context, this study originatively combined AHP and SAW to build a GIS-based MCE framework for mapping the basin-scale flood hazard. As the most famous and popular multi-criteria decision evaluation methods, AHP is usable for simplifying the assessment of complicated, unorganized and multi-attribute problems, driving criteria weights by using subjective judgments and empirical data, and allowing the calculation of the consistency for the evaluation procedure [18]. SAW is a proportional linear transformation of the raw data, which keeps the relative order of magnitude of the standardized scores equal. The SAW-based flood hazard index is calculated by summing up all alternative scores which can be derived by multiplying the normalized criteria weights by the standardized rating value of each corresponding criterion [19]. It is worth noting that the total weight must be 1 during the weighting, which makes it a difficulty that assigning the right weight in terms of criterion priorities. While the AHP, assigning and validating weights by comparing the importance degree of each criterion and calculating the consistency ratio, can meet the requirement. Therefore, the combination of AHP and SAW is reasonable. The results of the case study indicate the availability of the combination of AHP and SAW in the Dadu River basin.

#### 4. Conclusions

The estimation and mapping of the spatial flood hazard distribution in a GIS environment are essential for impactful flood risk management. In the study, the flood hazard in the Dadu River basin is identified based on an AHP-SAW method and the GIS tool. Specifically, the information of hydrometeorology, topography, drainage, and land usage aspects are weighted by the AHP method and integrated by the SAW method. The flood-prone areas in the basin are divided into five levels from lowest to highest based on the calculated FHI values. The results are fundamental for formulating effective flood risk management planning and measures. The combined method of AHP and SAW can be easily extended to other river basins if the required indicator data can be collected, and other disaster fields, such as landslide, erosion and fire. Present researches in the field mostly used static variables, and future studies can explore how to mix dynamic variables, such as climate and land use change data, with static variables to map the flood hazard, which can be used as scientific support of real-time flood risk decision-making.

#### References

- [1] Jiang WG, Deng L, Chen LY, Wu JJ and Li J. Risk assessment and validation of flood disaster based on fuzzy mathematics. *Progress in Natural Science*. 2009; 19:1419-1425.
- [2] Chen Y, Liu R, Barrett D, Lei G, Zhou MW, Renzullo L and Emelyanova I. A spatial assessment framework for evaluating flood risk under extreme climates. *Science of the Total Environment*. 2015;

- 538: 512-523.
- [3] Jonkman SN, Kok N and Vrijling JK. Flood risk assessment in the Netherlands: a case study for dike ring South Holland. *Risk Analysis*. 2008; 28:1357-1374.
  - [4] Meyer V, Scheuer S and Haase D. A multi-criteria approach for flood risk mapping exemplified at the Mulde river, Germany. *Natural Hazards*. 2009; 48:17-39.
  - [5] Wang Y, Li Z and Tang Z. A GIS-based spatial multi-criteria approach for flood risk assessment in the Dongting Lake Region, Hunan, Central China. *Water Resources Management*. 2015; 25: 3465-3484.
  - [6] Papaoianou G, Vasiliades L and Loukas A. Multi-criteria analysis framework for potential flood prone areas mapping. *Water Resources Management*. 2015; 29:399-418.
  - [7] de Brito MM, Evers M. Multi-criteria decision-making for flood risk management: a survey of the current state of the art. *Natural Hazards and Earth System Sciences*. 2016; 16:1019-1033.
  - [8] Malczewski J and Rinner C. Multicriteria decision analysis in geographic information science. Springer Berlin, 2015.
  - [9] Çağıl K, Vedat D, Can A and Mehmet LS. Preparation of a geotechnical microzonation model using Geographical Information Systems based on Multicriteria Decision Analysis. *Engineering Geology*. 2006; 87:241-255.
  - [10] Kandilioti G and Makropoulos C. Preliminary flood risk assessment: the case of Athens. *Natural Hazards*. 2012; 61: 441-468.
  - [11] Dewan A. Floods in a megacity: geospatial techniques in assessing hazards, risk and vulnerability. Springer, Dordrecht, 2013.
  - [12] Fernández DS and Lutz MA. Urban flood hazard zoning in Tucumán Province, Argentina, using GIS and multicriteria decision analysis. *Engineering Geology*. 2010; 111: 90-98.
  - [13] Nerantzis K, Ioannis K and Thomas P. Assessment of flood hazard areas at a regional scale using an index-based approach and Analytical Hierarchy Process: Application in Rhodope-Evros region, Greece. *Science of the Total Environment*. 2015; 538:555-563.
  - [14] Kourialis NN and Karatzas GP. A national scale flood hazard mapping methodology: The case of Greece – Protection and adaptation policy approaches. *Science of the Total Environment*. 2017; 601-602: 441-452.
  - [15] Hu SS, Cheng XJ, Zhou DM and Zhang H. GIS-based flood risk assessment in suburban areas: a case study of the Fangshan District, Beijing. *Natural Hazards*. 2017; 87: 1525-1543.
  - [16] Ahmed EMA, Ayman MN and Omar EMA. Evaluation of flood susceptibility mapping using logistic regression and GIS conditioning factors. *Arabian Journal of Geosciences*. 2018; 11:765.
  - [17] Saaty TL. A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*. 1977; 15: 234-281.
  - [18] Sedghiyan D, Ashouri A, Maftouni N, Xiong QG, Rezaee E and Sadeghi S. Prioritization of renewable energy resources in five climate zones in Iran using AHP, hybrid AHP-TOPSIS and AHP-SAW methods. *Sustainable Energy Technologies and Assessments*. 2011; 44:101045.
  - [19] Chabuk AJ, Al-Ansari N, Hussain HM, Sven S and Roland P. GIS-based assessment of combined AHP and SAW methods for selecting suitable sites for landfill in Al-Musayib Qadhaa, Babylon, Iraq. *Environmental Earth Sciences*. 2017; 76:209.

# Study on Cross-Basin Ecological Compensation Mechanism and Compensation Standard

Jinren LUO, Yizhi SUN<sup>1</sup>, Yihu ZHAO, Qinyao FU, Jiayin LI

*Institute of Economics and Management, Lanzhou Jiaotong University, Lanzhou,  
Gansu, 730070, China*

**Abstract.** Water resources in China's river basins are scarce, and the pollution that shrouds them is serious. Constant disputes have emerged between the upstream and downstream sectors due to the contamination of river basins. Moreover, China's research on ecological compensation mechanisms and compensation standards is still immature at present. Thus, this study establishes a compensation model and introduces the compensation coefficient  $K$ , including the compensation coefficient  $K_1$  between the upstream and downstream governments and the compensation coefficient  $K_2$  between the upstream government and the central government. This paper adopts the Bargain Game Model and obtains the value of  $K_2$  through the decision-making process between the central government and the upstream local government. In addition, amendment to the final offer arbitration law is used to acquire the value of  $K_1$  by proving the existence and uniqueness of equilibrium. Then, this paper takes the Taohe River Basin as an example and combines the compensation model to analyze, using the simplified compensation function to determine the amount of emission pollution from upstream to downstream and the compensation that upstream should receive.

**Keywords.** Ecological compensation mechanism, game model, compensation coefficient, final price arbitration method, compensation standard

## 1. Introduction

With the rapid economic development of China, the problem of water pollution across river basins has increasingly become an obstacle to social development. Cross-basin water pollution is a kind of trans-boundary externality. Affected by the natural integrity and mobility of the watershed, pollution in a certain area can usually be transferred to another area through the water body, which leads to an imbalance of interests among different areas. Distorted economic and environmental interests have made the environmental protection of China's cross-basin water resources undergo several contradictions and difficulties. These conflicts have also threatened the fair and harmonious development among regions. On this basis, adjusting the benefits

<sup>1</sup> Corresponding Author: Yizhi Sun, Institute of Economics and Management, Lanzhou Jiaotong University, Lanzhou, Gansu, 730070, China; E-mail: [sunyizhi0108@163.com](mailto:sunyizhi0108@163.com)

This work was supported by the National Natural Science Foundation of China, "Benefit Balance Mechanism of Stakeholders with Multi-objective Trans-water Project Between Less Developed Regions" (grant numbers 71463036).

distribution relationship of relevant stakeholders and promoting the coordinated development of the economy have become academic issues worth studying.

In terms of ecological compensation mechanism, scholars are more inclined to regard it as an economic incentive mechanism for market compensation [1]. The purpose of the ecological compensation is to protect and restore the water and ecological resources that are influenced by human activities and to maintain ecosystem functions provided by the water resource environment [2-3]. James (2005) [4] believed that if the upstream area causes pollution to the downstream area, the upstream area must compensate for the loss of the downstream area due to its pollution; conversely, if the upstream area provides good ecological services to the downstream area and invests in protection cost, the downstream area should give certain compensation to the upstream area. Wunder et al. (2008) [1] posited that the most ideal ecological compensation should fully integrate ecological services into the market. Gong et al. (2010) [5] studied the role of social capital in ecological compensation. They claimed that social capital plays a key role in the success of ecological compensation. Only with strong social capital as a guarantee can stakeholders be promoted in ecological compensation. The selection of behaviors in the compensation mechanism that is conducive to environmental protection will not cause harm to society. Hecken (2010) [6] asserted that ecological compensation can produce positive benefits for society only when it is proven to be reasonable. Therefore, they conducted a deep analysis of whether ecological compensation is reasonable from the perspective of political science. They concluded that ecological compensation is indeed reasonable and effective. Many Chinese scholars study ecological compensation in river basins from the perspective of game theory. Liang (2007) [7] maintained that the ecological problems of the river basin are the result of the individual rationality of the stakeholders in the river basin. To achieve collective rationality, a selective incentive mechanism for ecological compensation in the river basin can be established through the incentive mechanism of upstream ecological protection and the downstream ecological compensation force mechanism to resolve the contradiction between individual rationality and collective rationality in the basin. By studying the ecological compensation in Minjiang River, Han et al. (2009) [8] pointed out that inter-regional ecological compensation in the river basin is a coordinated negotiation among governments representing the public interest in various regions. The researchers highlighted that the key to promote the government to shift from non-cooperative game to cooperation game is the establishment of an incentive and restraint integration mechanism. Song (2009) [9] discussed the benefit distribution of ecological compensation stakeholders from the perspective of game theory and established a game model. The analysis showed that to achieve the optimal ecosystem, long-term ecological compensation goals should be initiated; moreover, a socialized supervision and evaluation mechanism should be established, and ecological property rights and evaluate ecological environmental value should be clearly defined. Cao and Jiang (2009) [10] created a game model for the governments of various regions in the river basin and specifically analyzed the decision-making process of the compensation subject and the object under the established compensation mechanism. Under the constraints of this mechanism, each region in the basin can establish an optimization model based on the local payment function to calculate the impact of pollution discharge and pollution transfer on the economic benefits of the basin. It can also maximize the economic benefits of the region by adjusting its own pollution discharge and pollution transfer volume, thereby improving the ecological environment of the area. Zhan (2016) [11] introduced the corresponding principles and implementation methods, which are based

on the increasingly prominent contradiction between China's current economic growth and environmental protection. Gao et al. (2019) [12] proposed the influencing factors of watershed ecological compensation must be understood from the perspective of how interactions occur among different governments. Moreover, the upstream and downstream governments cannot spontaneously cooperate to implement a watershed ecological compensation system without the supervision of the central government.

Looking at the existing research results of scholars at home and abroad, few game analyses on ecological compensation mechanism exist, and the research methods are too focused on theoretical analysis and quantitative model construction. The results are found in a few scholars, such as Cao and Jiang [10], who used repeated games to analyze compensation for cross-regional pollution. However, the model is too complex to be practically operated and empirically tested.

The criteria for determining the willingness to pay for ecological compensation depends on the respondents' awareness of the importance of the ecological environment and its protection [13]. In addition, it is difficult to reach a compensation agreement on the wishes of both the beneficiary and the supplier [14]. In terms of the compensation standard, Hamdar (1999) [15] used linear programming models to calculate the corresponding compensation amounts according to land use methods based on the investigation of land conversion projects along the Mississippi River. In this type of research, most scholars have used the willingness to pay method [16] and the evaluation of ecological service functions [17]. Li and Hu (2007) [18] applied the ecological reconstruction cost sharing method when studying the issue of ecological compensation standard in the upper and lower reaches of the Minjiang River basin. They pointed out that this method is suitable for determining the economic compensation standard of different river basins. Thus, it is the most appropriate standard accounting method for upstream and downstream ecological compensation at this stage. Jiang (2008) [19] used the water resource value method, cost analysis method, and opportunity cost method to calculate the ecological compensation standard with the water source protection area of the Han River basin as the research object. The researcher believed that the results obtained by using the water resource value method were more appropriate. Mao (2008) [20] adopted the willingness to pay method, cost analysis method, and opportunity cost method to study ecological compensation standard. The researcher pointed out that using the cost method as a quantitative standard when compensating for inter-basin water transfer sources can truly reflect the value of ecological protection. Yuan and Wu (2016) [21] determined the ecological pollution compensation standard of China's industrial sectors as the research object, quantified the upper and lower limits of compensation for heavy and non-heavy polluting industries from the perspective of environmental costs, and used the game to narrow the range gradually and determine the specific compensation standard. Chang (2016) [22] used the willingness to pay method and the opportunity cost method to calculate the ecological compensation standard for the water conservation area in the upper reaches of the Fenhe Reservoir. The researcher determined that the compensation standard ranges from 500 million to 1480 million. Guan et al. (2016) [23] took China's Xiaohong River Basin as an example. They used chemical oxygen demand (COD) as the main pollution evaluation index to calculate the ecological compensation value of the water environment from 2008 to 2012.

In summary, the calculation of the ecological compensation standard is still immature, and many methods are available. The results obtained by each approach are often very different, and the constant recognition of stakeholders in practical applications is difficult to obtain. Two main categories exist: the accounting method and the

negotiation method. The accounting method mainly includes the willingness to pay method, the opportunity cost method, the cost analysis method. The negotiation rule is a method to determine the standard of ecological compensation by negotiating a certain range of ecological compensation based on the accounting method and taking into account the willingness of the compensation subject and object, that is, the price negotiation game. Behind the free negotiation of ecological compensation standard is the economic game process between the beneficiaries and suppliers of ecosystem services. Free negotiation often has difficulty reaching a protection compensation agreement, and an effective negotiation and arbitration mechanism between the compensator and the recipient is required to facilitate stakeholders reach compensation agreements through limited consultations. Yuan and Wu (2016) [21] designed the compensation standard and compensation implementation mechanism. He constructed a multi-level ecological value compensation standard for economic loss-pollution accidents. Moreover, he believed that the polluter and the victim can achieve direct compensation through game. However, no specific demonstration exists for the implementation of the game.

The innovation of this study is to incorporate the central government into cross-regional ecological compensation, and use different bargaining game models to determine different ecological compensation coefficients and then determine the compensation standard.

## **2. Model Establishment of Upstream Local Government Compensation Mechanism**

### *2.1. Model Assumptions*

The upstream government and the downstream government in the river basin, and the central government are all rational economic people, all adopting strategic behaviors to maximize their own interests.

The upstream government, the central government, and the downstream government negotiate with one another on the discharge and treatment of sewage.

The central government compensates the upstream government to urge the upstream and the downstream to follow the principle of “upstream protects, downstream compensates; upstream doesn’t protect, upstream compensates” to maximize the social welfare of the whole society.

### *2.2. Establishment of the model*

#### *2.2.1. Variable setting.*

P is the amount of pollutants discharged in the upstream area;

$P_0$  is the maximum amount of pollution allowed to be discharged in the upstream area;

T is the amount of pollutants transferred from the upstream area to the downstream area;

R is the revenue function generated in the upstream area, which is regarded as a function of P;

C is the cost function of reducing pollution in the upstream area;

A(T) is the economic impact of upstream’s pollutants area on the downstream area;

$M$  is the total amount of compensation that the upstream obtains from the downstream government and the central government;

$K_1$  is the compensation coefficient of the downstream area to the upstream area. When the upstream causes less pollution to the downstream,  $K_1$  is a positive number; when the upstream causes more pollution to the downstream,  $K_1$  is a negative number; when the upstream is neither penalized nor compensated,  $K_1=0$ .

$K_2$  is the compensation coefficient of the central government to the upstream government. When the upstream causes less pollution to the downstream,  $K_2$  is a positive number; when the upstream causes more pollution to the downstream,  $K_2$  is a negative number; when the upstream is neither penalized nor compensated,  $K_2=0$ .

### 2.2.2. The Revenue Function and Its Analysis in The Upstream Area

The strategy of the upstream region is to transfer pollutants to the downstream area and obtain compensation from the central government and the downstream local government to maximize utility.

The utility function of the upstream area is:

$$U(P, T) = R(P) - C(P - T) + A(T)K_1 + A(T)K_2. \quad (1)$$

Formula (1) satisfies the constraint conditions:  $0 \leq T \leq P \leq P_0$ . Given that the revenue function is continuous, both upstream and downstream parties want to maximize their benefits. The total revenue function and its derivative exist and are continuous. Therefore, the first derivative of the revenue function is equal to 0 and is a necessary condition for maximizing income.

Given the compensation coefficients of the central government and the downstream government, the upstream area must take the pollutant emission decision  $a(P, T)$  that

$$\frac{\partial U}{\partial P} = 0, \frac{\partial U}{\partial T} = 0$$

maximizes their own benefits, that is,  $\frac{\partial U}{\partial P} = 0, \frac{\partial U}{\partial T} = 0$ . Thus, the following equations are obtained:

$$\frac{\partial U}{\partial P} = R'(P) - C'(P - T) = 0, \quad (2)$$

$$\frac{\partial U}{\partial T} = C'(P - T) + A'(T)[K_1 + K_2] = 0. \quad (3)$$

After the compensation coefficients  $K_1$  and  $K_2$  are determined, the upstream decision vector  $a(P, T)$  can be obtained according to the equation group to obtain the final income of the upstream area.

### 2.2.3. The Compensation Standard Function and Analysis of The Upstream Area

The compensation standard function in the upstream area is:

$$M = A(T)K_1 + A(T)K_2 = A(T)(K_1 + K_2). \quad (4)$$

After determining the compensation coefficients  $K_1$  and  $K_2$ , the upstream decision vector  $a(P, T)$  can be obtained according to the equations, and the compensation amount of upstream can be determined.

### 3. The Game Between the Upstream Local Government and The Central Government— $K_2$ Solution

#### 3.1. Game Method

The determination of the compensation coefficient of the central government is obtained by the bargaining game between the upstream government and the central government. First, the central government proposes a compensation coefficient for the amount of polluted water that must be transferred to the downstream. Then, the upstream government can accept or reject it. If it rejects the coefficient, it will propose its own asking price, and so on. In the whole game process, as long as one party accepts the opponent's plan, the game is over.

#### 3.2. Precondition

The time value of compensation and fixed cost cannot be ignored in the process of bargaining negotiation. One more round of negotiation entails increases in the time cost and the fixed cost.

In the model construction, the number of bargaining times is set as an odd number. Three unsuccessful negotiations require an inverse solution five times. To simplify the model and increase the practical operation feasibility, the number of games is limited. The game is assumed to have three stages at most. The upstream government has to accept the plan in the third round of the game.

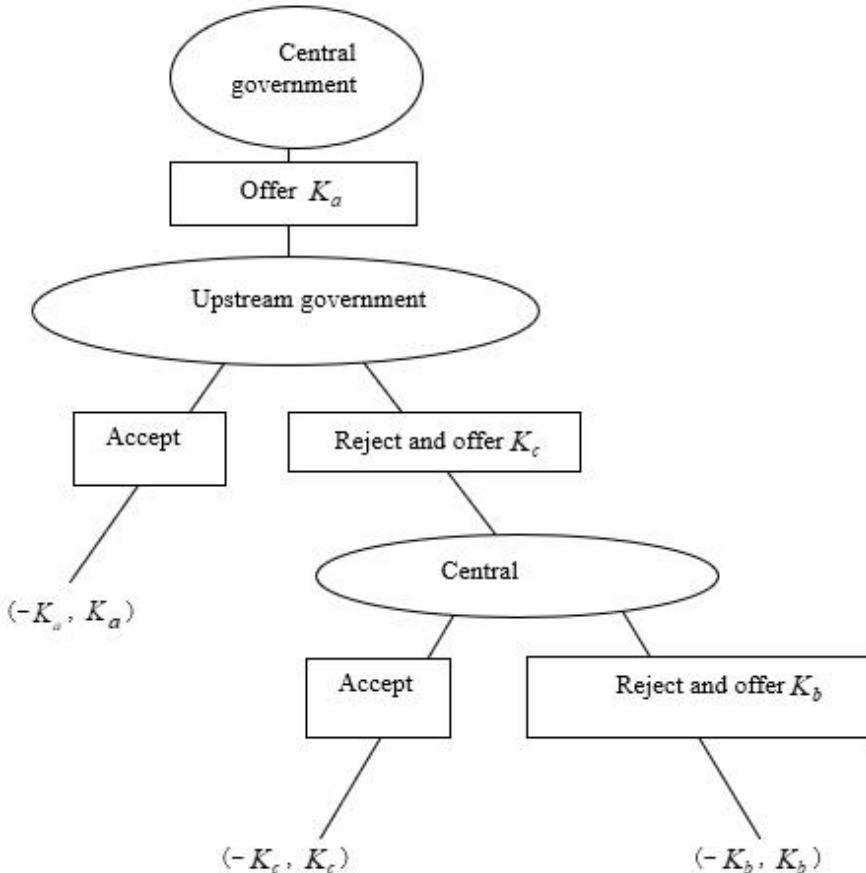
Two effects can emerge from the upstream to the downstream according to the amount of sewage: if the amount of transferred sewage is small, the downstream gains benefits, and the central government needs to compensate the upstream; conversely, if the downstream suffers from the sewage, then the central government should punish the upstream government. In this case, the compensation coefficient is negative, that is, the upstream government needs to pay a fine to the central government. Assuming that the amount of sewage transferred from the upstream to the downstream is small and downstream development gains benefits, the central government needs to provide additional compensation to the upstream government.

#### 3.3. Game process

The game process of the compensation coefficient  $K_2$  between the upstream government and the central government is shown (Figure 1). In the first stage, when less sewage  $T$  is transferred from the upstream to the downstream, the central government's plan is to compensate the upstream with  $K_a$ . Then, the upstream government makes the decision to accept it or not. The negotiation between the two levels of government will end if the upstream accepts. The compensation will be implemented in accordance with the plan of the central government. At the same time, the upstream government will not have made any progress. If the upstream government does not accept it, the next phase will begin. In the second stage, the compensation coefficient requested by the upstream government is  $K_c$  ( $K_c > K_a$ ), and the central government chooses whether to accept it. If accepted, the upstream government will obtain the compensation coefficient  $K_c$ , and  $K_c - K_a$  is the increment. If the central government does not accept it, then the process moves to the

next stage. In the third stage, the compensation coefficient proposed by the central government is  $K_b$  ( $K_c > K_b > K_a$ ), which must be accepted by the upstream. The numbers in the above three stages are all real numbers greater than 0.

The key point of this game is that the plan of the central government in the third stage is mandatory, that is, the compensation coefficient proposed by the central government to the upstream at this stage must be accepted by both parties. Every time this game stage is played once more, it is detrimental to the work. Ending the debate on the compensation coefficient as soon as possible is especially beneficial to the upstream government.



**Figure 1.** Game process diagram between the upstream government and the central government.

### 3.4. Game Analysis

The analysis of this game by inverse induction is as follows: in the third stage of this game, the central government proposes the compensation coefficient of the upstream as  $K_b$ , and the upstream government must accept it. In the second stage, the upstream

government knows that once the third stage is reached, the central government will insist that the compensation coefficient it receives  $K_b$ , which has increased to  $K_b - K_a$ . To maximize its own benefits, if the upstream government rejects the central government's plan in the first stage, it must propose a plan greater than  $K_b$ . If the delivered value proposed by the upstream government, which makes the central government transfers greater benefits than planned in the third stage, then the plan will definitely be rejected by the central government. The analysis must proceed to the third stage and the upstream government obtains the benefits accepted in this stage. Therefore, the plan is deemed the ideal choice if it makes the central government willing to accept and its own benefits can be as large as possible in the third stage. In other words, in the game, either player is willing to accept a bid at this stage, which is no less than that of their own at the next stage. When the upstream proposes the target compensation coefficient  $K_c$ , its compensation coefficient is increased to  $K_c - K_a$ . This increase is larger than  $K_b - K_a$ , which is the biggest benefit that the upstream may obtain. Then, inverse deduction is made to the analysis of goals proposed by the central government in first stage. Evidently, the central government knew from the beginning that it would transfer the benefits of  $K_b - K_a$  to the third stage. It is also aware that the compensation coefficient proposed by the upstream in the second stage is the target of  $K_c$ . Therefore, the benefits it is willing to transfer are also  $K_b - K_a$ , while the upstream is satisfied with the greatest possible benefit, that is,  $K_c - K_a$ . For this reason, the game must continue from the first stage to the third stage. One point worth explaining in this game is that the above conclusions have a premise that the transfer of interests proposed by the central government in the third stage must be known in advance by both parties. In this game, the central government in a favorable position can guarantee all its own interests without transferring interests only when the central government is not concerned about protracted negotiations at all or when the upstream government's request is unreasonable. In this game of bargaining between the central government and the upstream local government, the central government has two roles, namely, the "participant" (economic man) and the "manager" (administrative person), due to the unequal status of the central government and the upstream government. The central government conducts transactions with a dual identity, which is evidently mandatory. This coercion restricts the game space between governments.

#### 4. The Game Between Upstream and Downstream Local Government— $K_1$ Solution

##### 4.1. Game Method and Process

The upstream and downstream governments of the Taohe River basin are based on the relationship of "brothers," and they have equal status. The game between these two parties on the compensation coefficient is widely different from the game with the central government, and it generally requires a third party to adjust the arbitration. Farber first proposed the final offer arbitration law. On this basis, Dao-Zhi Zeng (2006)

[24] proposed the amendment final offer arbitration (AFOA) law, on which the following is based to solve the upstream and downstream compensation coefficients.

According to the final offer arbitration law, the final arbitration result is determined by the bid winner (that is, the party closer to the ideal value of the arbitration department). According to the AFOA law, the final arbitration price is based on the failure of the bid the party (that is, the party farther away from the ideal value of the arbitration department) determines as the final arbitration price. The rule is that upstream  $i$  and downstream  $j$  first offer a price at the same time, assuming that downstream bid is  $K_j$  and the upstream asking price is  $K_i$ . Both parties notify the arbitration department, and the arbitration department obtains  $Z$ , the idea value, on the basis of scientific analysis and calculation.

To reach the optimal benefit for both parties from the game, the amount of sewage discharged from upstream to downstream is less, and the downstream is the compensator. Thus, when  $K_i \leq K_j$ , the two parties reach an agreement that the final arbitration value is  $\frac{K_i + K_j}{2}$ . When  $K_i \geq K_j$  and  $|Z - K_i| < |Z - K_j|$ , the bid of downstream  $j$  is farther from arbitration  $Z$ . Hence, the downstream wins during the arbitration process, and the final arbitration value is  $2Z - K_j$ . When  $K_i \geq K_j$  and  $|Z - K_i| > |Z - K_j|$ , the asking price of upstream  $i$  is farther to arbitration value  $Z$ . Therefore, the upstream wins during the arbitration process. The final arbitration value is  $2Z - K_i$ . When  $K_i \geq K_j$  and  $|Z - K_i| = |Z - K_j|$ , the final arbitration value is  $\frac{K_i + K_j}{2}$ .

#### 4.2. Proof of Game Feasibility

Dao-Zhi Zeng (2006) [24] proved that in the process of using the AFOA for arbitration, a unique pure strategic Nash equilibrium is reached; the equilibrium value is equal to the ideal value of the arbitration department.  $Z$ ,  $K_i$ , and  $K_j$  are assumed to be random variables and are independent of one another.

##### 4.2.1. Proof of Equilibrium Existence

In this game, a pure strategic Nash equilibrium is assumed to exist, that is, both parties bid  $E$ .

Given that the upstream asking price is  $K_i$ , the downstream bid is  $K_j$ . According to the arbitration rules, the solution is as follows:

$$a(K_i, K_j) = \begin{cases} \frac{K_i + K_j}{2}, K_j \geq K_i; \\ 2Z - K_i, K_j < K_i, \text{ and } Z > \frac{K_i + K_j}{2}; \\ 2Z - K_j, K_j < K_i, \text{ and } Z < \frac{K_i + K_j}{2}; \\ \frac{K_i + K_j}{2}, K_j < K_i, \text{ and } Z = \frac{K_i + K_j}{2}. \end{cases} \quad (5)$$

If  $K_j < K_i$ , the expected value of the arbitration value  $A(K_i, K_j)$  can be calculated as:

$$\begin{aligned} A(K_i, K_j) &= \frac{\int_{\frac{K_i+K_j}{2}}^{\infty} (2Z - k_j) dF(z) + \int_{-\infty}^{\frac{K_i+K_j}{2}} (2Z - k_i) dF(z) + \frac{k_i+k_j}{2} [F(\frac{k_i+k_j}{2}) - F(\frac{k_i+k_j}{2} - 0)]}{\int_{\frac{K_i+K_j}{2}}^{\infty} dF(z) + \int_{-\infty}^{\frac{K_i+K_j}{2}} dF(z)} \\ &= 2E - k_i + k_i F(\frac{k_i+k_j}{2}) - k_j F(\frac{k_i+k_j}{2} - 0) - \frac{k_i+k_j}{2} [F(\frac{k_i+k_j}{2}) - F(\frac{k_i+k_j}{2} - 0)] \\ &= 2E - k_i + \frac{k_i-k_j}{2} [F(\frac{k_i+k_j}{2}) - F(\frac{k_i+k_j}{2} - 0)]. \end{aligned} \quad (6)$$

When the asking price of upstream  $K_i = E$  if the bid of downstream  $K_j > E$ ,

$$\begin{aligned} A(E, K_j) &= E + \left( \frac{E - K_j}{2} \right) \left[ F\left(\frac{E + K_j}{2}\right) - F\left(\frac{E + K_j}{2} - 0\right) \right] \\ &\leq E + (E - K_j) \frac{F(E) - F(E - 0)}{2} < E. \end{aligned} \quad (7)$$

Similarly, when  $K_j < E$ ,  $A(E, K_j) = \frac{E + k_j}{2} < E$  only if  $K_j = E$  and  $A(E, E) = E$ .

Therefore,  $K_j = E$  is a strategy to maximize expectations, that is,  $K_i^* = K_j^* = E$  is a pure strategy Nash equilibrium.

#### 4.2.2. Proof of Equilibrium Uniqueness

$(K_i^\#, K_j^\#)$  is a pure strategic Nash equilibrium formed between the upstream and downstream local governments in the watershed ecological compensation if  $K_j^\# < K_i^\#$ . When the upstream increases its asking price, it will be compensated higher than the original asking price. If  $K_j^\# > K_i^\#$ , let  $\varepsilon = K_j^\# - K_i^\# > 0$ , it is discussed in two situations:

1) If  $F(K_j^\# - \frac{\varepsilon}{2}) = F(K_i^\# + \frac{\varepsilon}{2}) > 0$ , then:

$$\frac{\varepsilon}{2}F(K_i^\# + \frac{\varepsilon}{2}) \leq \frac{\varepsilon}{4}F(K_i^\# + \frac{\varepsilon}{2}) + \frac{\varepsilon}{4}F(K_i^\# + \frac{\varepsilon}{4}) < \frac{\varepsilon}{4}F(K_i^\# + \frac{\varepsilon}{4}), \quad (8)$$

$$\frac{\varepsilon}{2}F(K_i^\# + \frac{\varepsilon}{2} - 0) \leq \frac{\varepsilon}{4}F(K_i^\# + \frac{\varepsilon}{4}). \quad (9)$$

Thus,

$$\begin{aligned} A(K_i^\#, K_i^\# + \frac{\varepsilon}{2}) &= 2E - K_i^\# - \frac{\varepsilon}{4}[F(K_i^\# + \frac{\varepsilon}{4}) - F(K_i^\# + \frac{\varepsilon}{2} - 0)] \\ &< 2E - K_i^\# - \frac{\varepsilon}{2}[F(K_i^\# + \frac{\varepsilon}{2}) - F(K_i^\# + \frac{\varepsilon}{4} - 0)] \\ &= A(K_i^\#, K_i^\# + \varepsilon) \end{aligned} \quad (10)$$

$K_j^\# = K_i^\# + \varepsilon$  is the best strategy for downstream  $j$ .

2) If  $F(K_j^\# - \frac{\varepsilon}{2}) = 0$ , then  $F(K_j^\# - \frac{\varepsilon}{2} - 0) = 0$ . Thus:

$$F(K_j^\# - \frac{\varepsilon}{2}) - \frac{1}{2}F(K_j^\# - \frac{\varepsilon}{4}) + F(K_j^\# - \frac{\varepsilon}{2} - 0) - \frac{1}{2}F(K_j^\# - \frac{\varepsilon}{4} - 0) \leq 0 < 1, \quad (11)$$

and obtain:

$$\begin{aligned} A(K_j^\# - \varepsilon, K_j^\#) &= 2E - K_j^\# + \varepsilon - \frac{\varepsilon}{2}[F(K_j^\# - \frac{\varepsilon}{2}) + F(K_j^\# - \frac{\varepsilon}{4} - 0)] \\ &< 2E - K_j^\# + \frac{\varepsilon}{2} - \frac{\varepsilon}{4}[F(K_j^\# - \frac{\varepsilon}{4}) + F(K_j^\# - \frac{\varepsilon}{4} - 0)] \\ &= A(K_j^\# - \frac{\varepsilon}{2}, K_j^\#) \end{aligned} \quad (12)$$

Therefore,  $K_i^\# = K_j^\# - \varepsilon$  is the best strategy for upstream  $i$ .

In summary,  $K_i^\# = K_j^\#$ . The final result that makes both parties agree is  $k$ . If  $K < E$ , then

$A(k, k - 0) = 2E - k < k = A(k, k)$ . Therefore, downstream  $j$  will not raise a bid lower than  $k$ .

Similarly, if  $K > E$ , then  $A(k + 0, k) = 2E - k < k = A(k, k)$ . Thus, upstream  $i$  will not give an asking price higher than  $K$ .

#### 4.2.3. Result Analysis

In summary,  $k$  is the only Nash equilibrium of this game. This equilibrium is:  $K_i^* = K_j^* = E$ .

Now, assuming that the expected value distribution of the arbiter obeys a normal distribution with mean  $m$  and variance  $\delta^2$ , the density function is:

$$f(z) = \frac{1}{\sqrt{2\pi}\delta^2} e^{-\frac{1}{2\delta^2}(z-m)^2} \quad (13)$$

Therefore, the only Nash equilibrium point is  $K_i^* = m, K_j^* = m$ , that is, the expectation of the final bid is  $m$ .

### 5. The Determination of The Economic Impact Function of The Upstream Region on Downstream

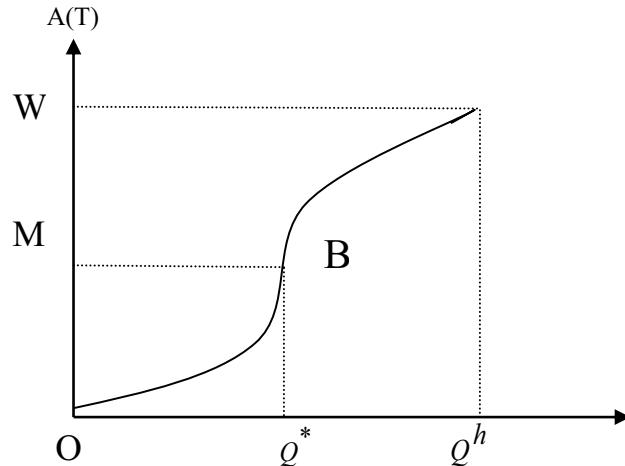
The economic impact function of the upstream area on the downstream, that is, the specific form of the compensation function, adopts the simplified form proposed by Li and Xu (2003) [25]:

$$A(T) = GDP * W * \frac{a(T - Q^*) - 1}{(e^{a(T - Q^*) + 1} + M)} \quad (14)$$

The meaning of each parameter in the formula is as follows:  $GDP$  represents the local GDP value of the current year.  $W$  represents the loss rate of water quality on the socio-economic impact. Meanwhile,  $a$  is expressed as a dimensionless coefficient, an important parameter that characterizes the shape of the curve. It reflects the sensitivity of pollution to the socio-economic impact: the larger the value, the steeper the function curve, thus indicating that the social economy is extremely sensitive to water quality. In contrast, it shows that economic behavior is less sensitive to water quality.  $T$  and  $Q^*$ , with no unit, respectively, represent the pollution transfer level and the water quality level corresponding to the inflection point. The value is converted according to a certain environmental protection standard for the actual discharge and transfer of pollutants.  $M$  is the influence coefficient of the corresponding water pollution loss at the turning point, that is, the turning point of the impact of water quality level on economic loss.

The schematic diagram (Figure 2) and function characteristics of the loss measurement model are as follows: ① Upper and lower limits: when pollutants are reduced to a certain level, the quality of the water environment will not cause economic losses. Conversely, when pollutants increase and the water quality deteriorates to a certain extent, the water body basically loses its due service functions. In addition, the economic loss rate caused by water pollution tends to maintain a constant state and reach the maximum value, assuming it is equal to 1. ② Non-linearity and inflection point: the relationship between water quality and economic impact should be a non-linear continuous gradual process, and the inflection point corresponding to the economic loss rate of water pollution  $M$  is assumed to be 0.5. ③ The growth rate of economic losses

usually shows a downward trend and gradually tends to be constant, thus reaching the upper limit of losses caused by water pollution.



**Figure 2.** The rivers pollution-economic loss function diagram.

## 6. Examples and Conclusions

As for the income function of the upper Taohe River area and the cost function of pollution control, the author directly draws on the research results of Cao and Jiang (2009) [10] based on the analysis of the ecological compensation mechanism of cross-basin pollution, namely:

$$R(P) = -2P^2 + 16P, \quad (15)$$

$$C(P - T) = 0.075(P - T)^2 + 0.001(P - T) \quad (16)$$

Given that many areas are involved in the water-receiving areas in the downstream of the Taohe River, for the sake of calculation, the author selects six representative regions to calculate the GDP of the downstream water-receiving areas: Anding, Longxi, Weiyuan, Lintao, Yuzhong, and Huining. According to Luo's calculation (2012) [26] of the data table of the carrying capacity of Tao River's water resources (Table 1), the GDP of the water receiving area is calculated as 2,633 million yuan by average. In addition, set  $W=0.7$ ,  $a=0.54$ ,  $Q^*=4$ , and  $M=0.5$ . Therefore, the total income function of the upstream region is:

$$U(P, T) = -2P^2 + 16P - 0.075(P - T)^2 - 0.001(P - T) + (K_1 + K_2) * 263300 * 0.7 * \frac{0.54(T - 4) - 1}{e^{0.54(T - 4) + 1} + 0.5} \quad (17)$$

**Table 1.** Data table of GDP and the value of the three industries in the water-receiving areas of the Yintao project<sup>a</sup>

County	GDP	Primary industry value (10,000 Yuan)	Second industry value (10,000 Yuan)	Third industry value (10,000 Yuan)
Anding	193573.16	56956.62	47672.05	88944.50
Longxi	190385.25	54441.00	60573.25	75371.00
Weiyuan	85948.75	45568.50	6610.25	33770.00
Lintao	167524.00	63356.50	48901.50	55266.00
Yuzhong	749148.75	57132.25	114664.75	63162.00
Huineng	193220.50	65371.50	49334.00	78515.00

<sup>a</sup>Data source: Luo Jinren's "Tao Water Diversion Project and Its Impact on Regional Sustainable Development" compiled from Appendix 8-7.

**Table 2.** Statistical properties of the used data.

Statistical Property	GDP	Primary Industry Value (10,000 Yuan)	Second Industry Value (10,000 Yuan)	Third Industry Value (10,000 Yuan)
Mean	263300.0683	57137.72833	54625.96667	65838.08333
Maximum	749148.75	65371.5	114664.75	88944.5
Minimum	85948.75	45568.5	6610.25	33770

The statistical properties of the data are summarized in Table 2. According to the previous argument in this article, suppose that the compensation coefficient obtained through arbitration in the upstream and the downstream of the Taohe River is  $K_1=0.5$ . In addition, the compensation coefficient obtained by the upstream government and the central government through the game is  $K_2=0.7$ . The utility function of the upstream area is continuous and second-order derivable. Through calculation, the upstream area selects the emission pollution level of 5.3 and transfers the pollution level of 1.39 to the downstream area. According to the five-level water quality conversion standard, the annual transfer of 550 million cubic meters of water from the upstream of the Tao River to its downstream, the amount of pollutants transferred from the upstream to the downstream can be calculated as  $T'=5.5*15*10^8 = 8250$  (tons).

Generally speaking, sewage treatment costs are relatively abstract and cannot be traded through the market. Hence, they are difficult to measure by market prices. Regarding the calculation of sewage treatment costs, many scholars have analyzed it from different angles, and the research results also vary widely. Therefore, this study directly adopts Zhou's (2009) [27] compensation standard on the basis of the compensation factor method as 15,000 yuan/ton.

Therefore, the compensation amount in the upstream area is calculated as:

$M = T' * P * \delta = 8250 * 1.5 * 1 = 12375$  (ten thousand yuan). At this time, the upstream of the Taohe River can obtain the maximum utility.

Evidently, the central government's intervention is highly necessary for the ecological compensation between the upstream and downstream parties of the river basin. The participation of the central government ensures the normal progress of the ecological compensation game between the upstream government and the downstream government. In the issue of ecological compensation in the upper and lower reaches of the river basin, the contradiction in ecological compensation caused by the imbalance of interests between the upstream and downstream governments of the basin will be eased to a large extent as long as a reasonable compensation mechanism is established and reasonable compensation standards are formulated.

## 7. Issues for Further Research

The study of cross-regional ecological compensation mechanism, on the one hand, involves the impact of the upstream pollution on the downstream economy. Testing such impact is a complicated issue. On the other hand, the bargaining game between the upstream and downstream governments and the central government involves the impact of pollution levels on the compensation coefficient, which makes the pollution coefficient itself endogenous. All of these aspects require further discussion.

## References

- [1] Boumans R, Costanza R, Farley J. Modeling the dynamics of the integrated earth system and the value of global ecosystem services using the GUMBO model. *Ecological Economics*. 2002; 41(3): 529–560.
- [2] Cao GH and Jiang DL. Ecological compensation solution to transboundary pollution. *Ecological Economy*. 2009; 11:160-164.
- [3] Chang S M. Study on the water ecological compensation standard of upstream of Fenhe reservoir. *Yellow River*. 2016; 38(9):56-58.
- [4] Dao ZZ. How powerful is arbitration procedure AFOA. *Int. Rev. Law and Economics*. 2016; 26(2):227-240.
- [5] Duan J, Yan Y, Wang D. Principle analysis and method improvement on cost calculation in watershed ecological compensation. *Acta Ecologica Sinica*. 2020; 30(1):0221-0227.
- [6] Dai JH, Wang HJ, Wang HL. An introduction to framework of assessment of the value of ecosystem services and natural capital. *Progress in Geography*. 2012; 387 253-260.
- [7] Gong Y, Bull G, Baylis K. Participation in the world's first clean development mechanism forest project: the role of property rights, social capital and contractual rules. *Ecological Economics*. 2010; 69(6):1292-1302.
- [8] Guan XJ, Liu WK and Chen MY. Study on the ecological compensation standard for river basin water environment based on total pollutants control. *Ecological Indicators*. 2016; 69: 446–452.
- [9] Gao S, Shen JQ, He WJ. An evolutionary game analysis of governments' decision-making behaviors and factors influencing watershed ecological compensation in China. *J. Environ. Manage*. 2019; 251(): 109592.
- [10] Hamdar B. An efficiency approach to managing Mississippi's marginal land based on conservation reserve program(CRP). *Resources, Conservation and Recycling*. 1999; 26(1):15-24.
- [11] Herzog F, Dreier S, Hofer G. Effect of ecological compensation areas on floristic and breeding bird diversity in Swiss agricultural landscapes. *Agriculture Ecosystem & Environment*. 2005; 108(3):189–204.
- [12] Han LF, Hu Y and Li YS. Analyzing ecological compensation mechanism in Minjiang River basin from the angle of game theory. *China Water Resources*. 2009; 6:10-12.
- [13] Hecken GV and Bastiaensen J. Payments for ecosystem services: justified or not? A political view. *Environ. Sci. Policy*. 2010; 13(8):785-792.

- [14] James S. Creating markets for ecosystem services: notes from the field. *New York University Law Review*. 2005; 80(6): 871-888.
- [15] Jiang ZW. Study on ecological compensation standard and mechanism of South-to-North water diversion in Hanjiang River water resource basin. Xian: Xi'an University of Architecture and Technology, 2008.
- [16] Li JX and Xu SL. Calculation model of water pollution induced economic loss for river basin. *J. Hydraul. Eng.* 2003; 2003(10):68-74.
- [17] Li YS and Hu Y. On regional ecological benefit compensation standard in Minjiang River basin. *Research of Agricultural Modernization*. 2007; 28 (3):327-329.
- [18] Liang LJ. Research on market-oriented operating system of eco-compensation in Drainage Areas—A case study of the Huanghe Drainage Area. Shandong: Shandong Agriculture University, 2007.
- [19] Luo JR. Water diversion project from Tao River and its influence on regional sustainable development. China social sciences press, Shanghai, 2012.
- [20] Mao ZF. Research on ecological compensation for water source area of inter-basin water transfer project —taking Ankang as an example. Shaanxi: Shaanxi Normal University, 2008.
- [21] Plantinga AJ, Alig R and Cheng H. The supply of land for conservation uses: evidence from the reservation reserve program. *Resources Conservation Recycling*. 2001; 31(3):199–215.
- [22] Rudd C, Canters KJ, Udo de Haes HA. Guidelines for ecological compensation associated with highways. *Biological Conservation*. 1990; 90(1):41–51.
- [23] Song M. Game analysis of establishing ecological compensation mechanism. *Academic Exchange*. 2009; 5:83-87.
- [24] Wunder S, Engel S, Pagiola S. Taking stock: A comparative analysis of payments for environmental services programs in developed and developing countries. *Ecological Economics*. 2008; 65(4):834-852.
- [25] Yuan GD and Wu. Research on the game mechanism of determining the ecological pollution compensation standards from the perspective of environmental cost. *Journal of Audit & Economics*. 2015; 12:65-73.
- [26] Zhou Y. Study on ecological compensation mechanism of water source area in the upper reaches of Huangpu River. Shanghai Jiaotong University, Shanghai, 2009.
- [27] Zhan HH. Study on the introduction of ecological compensation mechanism in environmental impact assessment. *Resources Economization & Environmental Protection*. 2016; 7:138-139.

# Social and Economic Value Evaluation of Water Resource Based on Fuzzy Mathematics and Emergy Theory

Xiang XU<sup>a,b</sup>, Yongqiang WANG<sup>a,1</sup>, Kai LI<sup>a</sup>, Junsong XIN<sup>b</sup>

<sup>a</sup> *Water Resources Department (Transboundary River Management Department), Changjiang River Scientific Research Institute, Wuhan 430010, China and Hubei Provincial Key Laboratory of Basin Water Resources and Ecological Environmental Sciences, Changjiang River Scientific Research Institute, Wuhan 430010, China*

<sup>b</sup> *College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China*

**Abstract:** Water resources is one of the important drivers of socio-economic development, but the value of water resources in society is not clear. In order to accurately describe the impact of water resources on socio-economic value, a socio-economic value evaluation index system for water resources is established. This paper is based on the theory of utility value of water resources. Discussed how to use fuzzy mathematics and benefit sharing coefficient method to calculate the socio-economic value of water resources in different industries. Take the Golog Zang A.P in the source region of the Yellow River as an example. Calculated the socio-economic value of water resources for residential life, irrigation planting, industry, construction and tertiary industry. Finally, analysis results show that the value of comprehensive water resources in the study area is between 9.4-40CNY, tertiary industry highest, lowest value for domestic water. The calculation results provide a reference for the rational and efficient use of regional water resources and the scientific formulation of water resources policies.

**Keywords.** Social and economic value, evaluation of water resource, entropy weight method, fuzzy comprehensive evaluation, emergy, benefit sharing coefficient

## 1. Introduction

Water is very important to our lives and society. It is an irreplaceable scarce economic resource for social and economic development, and also the most active and widely influential basic natural resource in the ecosystem. The purpose of this paper is to analyze the social and economic value of water resources for different industries in the region.

---

<sup>1</sup> Corresponding Author: Yongqiang Wang, Water Resources Department (Transboundary River Management Department), Changjiang River Scientific Research Institute, Wuhan 430010, China and Hubei Provincial Key Laboratory of Basin Water Resources and Ecological Environmental Sciences, Changjiang River Scientific Research Institute, Wuhan 430010, China; E-mail: [wangyongqiang1022@126.com](mailto:wangyongqiang1022@126.com)

This work is funded by the National Key R&D Program of China (2017YFC0403600, 2017YFC0403606); the National Natural Science Foundation of China (No.51779013, 51639005); the Water Conservancy Science and Technology Innovation project of the GuangDong Province (2017-03); National Public Research Institutes for Basic R&D Operating Expenses Special Project (No. CKSF2019478, No. CKSF2017008) Special thanks are given to the anonymous reviewers and editors for their constructive comments.

Establishing evaluation indicators and evaluation models for the socio-economic value of water resources. It can better reflect the actual situation of social and economic benefits of water resources, and provide a reference for the rational and efficient use of water resources.

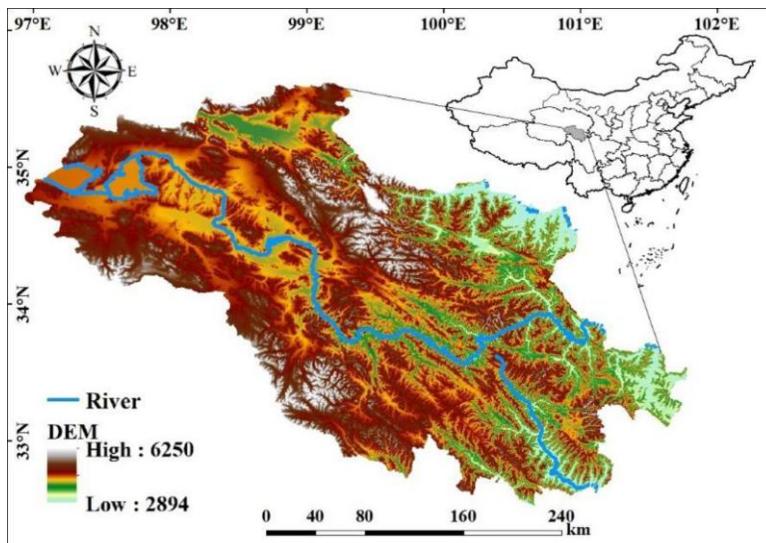


Figure 1. Study area.

To systematically analyze the socio-economic value of water resources. From figure 1. The research area is located in Golog Zang A.P of Qinghai Province, China, which belongs to arid and semi-arid areas in northwest China. By referring to past studies, we can clearly see that most of the study areas are located in economically developed areas, and the regional water resources crisis is more significant. The study of the value of water resources is imminent. However, we can find the current of social development in recent years. Although the social economy of the Northwest has improved, it has brought along a series of problems such as high consumption of water resources, low utilization of water resources, and serious water pollution. At this time, it is even more necessary for us to have a study on the value of local water resources, so that we can systematically realize that the water resources crisis is not a characteristic of economically developed regions, and to have a good understanding of the value of water resources in advance is helpful for us to achieve better sustainable development.

The Golog Zang A.P in Qinghai Province is located between the Bayan Har and Animaqing Mountains in the hinterland of the Qinghai-Tibet Plateau, in the east of the Three Rivers Source Area, Geographically located between 32°21'-35°45' north latitude and 96°56'-101°45' east longitude. The administrative divisions mainly include 6 counties, Maduo, Dari, Gande, Maqin, Jiuzhi and Banma, with a total area of  $7.6 \times 10^4 \text{ km}^2$ . The average altitude in the area is above 4200m, and the climate is a plateau alpine climate, with alternating hot and cold seasons and distinct dry and wet seasons. The average annual temperature is -4°C and the average annual rainfall is 400-700mm. Grassland is the main vegetation type in this area, accounting for 71% of the total land area. In 2019, the prefecture completed a regional GDP of 4,618.35 million yuan, a year-on-year increase of 7.2%, which was three times that of ten years ago. The state's per capita regional product

value was 2,1990 yuan, a year-on-year increase of 5.22%. The population accounts for more than 75% of the total population, and the urbanization rate is 28.02%.

## 2. Literature Review and Data Sources

Since the 1980s, great progress has been made in the research models and methods of water resources value at home and abroad. At present, the commonly used models include the following aspects: (1) Shadow price model. Zhu Jiulong and others[1] use linear programming model to estimate the theoretical value of water resources of different departments in 5 river sections of Huaihe River basin in 2000, taking the maximum net benefit of river basin water use as the objective function; proposed a method and a solution process for calculating the shadow price of water resources by using a nonlinear dynamic input-output optimization model, and calculated the shadow price of water resources in China's nine major watersheds from 1949 to 2050[2].(2) Fuzzy mathematics model. Jiang wenlai [3] estimated the water resources value of Beijing area in 1992 to be 1.041 m<sup>3</sup> by using fuzzy mathematics model, considering the influence of water quality After that, the value of water resources is 1.592m<sup>3</sup>;Miao Huiying and others[4] used fuzzy hierarchical mathematical model to make fuzzy comprehensive evaluation on the value of water resources in Shijiazhuang, and got the result that the value of water resources in that year was medium to high. Liu Dezhi used fuzzy comprehensive evaluation model to analyze the value of water resources in Hutuo River basin. Qiu Lin used fuzzy optimization neural network model to evaluate the value of water resources. (3) Residual value model. Renwick[5] estimated the value of Sri Lanka rice planting water as 0.93 Sri Lankan rupees (US\$ 1 =100 Sri Lankan rupees) using the residual value model. Glenn-Marie Lange[6] uses the residual value model in the Stampriet area of Namibia, and estimates the value range of unilateral agricultural water use in 1999 by assuming the rate of return on capital and the value range of land resources. The comprehensive evaluation methods of water resources value mainly include Emergy theory[7], matrix method[8], neural network method, comprehensive index method [9].

To sum up, we consider that the water resources system is a complex system, which is formed under the interaction, coupling and influence of natural, social and economic systems. There are many factors that affect the value of water resources. Each factor is not produced singly, but is interrelated. The value is the result of their joint action, which leads to the fuzziness of value composition and influencing factors. It is therefore unreasonable to consider the value of water resources from only one aspect. Based on the above reasons, we use fuzzy mathematics to calculate the value of water resources corresponding to residents' lives and use the energy value theory to calculate the benefit sharing coefficient, and then obtain the corresponding water industry value.

The socio-economic indicator data required for the study area comes from the National Economic and Development Bulletin of the autonomous prefecture, the government work report and the regional statistical year book (<http://tjj.qinghai.gov.cn/tjData/qhtjn/>;<http://tjj.qinghai.gov.cn/tjData/yearBulletin/>;<http://slt.qinghai.gov.cn/subject/list?cid=58>).

### 3. Value of Residents' Living Water Resources

#### 3.1. Fuzzy Comprehensive Evaluation Model

##### 3.1.1. Establishment of Social and Economic Benefit Evaluation Index System

The evaluation index system of social and economic benefits is an important basis for us to evaluate and judge the rationality of the selected scheme, so we should follow certain principles and standards when establishing the index system to ensure that it can correctly, scientifically, comprehensively and objectively reflect the effect of the evaluation model. The evaluation index system of social and economic benefits of Golog Zang A.P includes two levels: the factor level and the index level[10]. The factor level of this evaluation is divided into two categories of economic benefits ( $U_1$ ) and social benefits ( $U_2$ ). Each factor level is further divided into several index levels, as shown in Table 1 for details, thus establishing an index element set  $U = (U_1, U_2)$ .

**Table 1.** Composition of Index System

Target layer	Element layer	Indicator layer
Evaluation of Social and Economic Benefits	Economic benefits ( $U_1$ )	Per capita disposable income Output value of agriculture and animal husbandry Output Value of Chinese and Tibetan Medicinal Materials
	Social benefits ( $U_2$ )	Grassland area Grassland availability factor population rate of increase Number of people out of poverty

##### 3.1.2. Data Collection and Selection of Indicators

Through the analysis and research of the basic data of the Golog Zang A.P, we selected the indicators shown in Table 1. Among the economic benefits, we considered the three indicators of per capita disposable income(E1), agricultural and animal husbandry output(E2), and Chinese and Tibetan medicinal materials(E3). The reason is that Golog Zang A.P is located on the Qinghai-Tibet Plateau in China. The high altitude and low temperature are not conducive to crop growth. The added value of the primary industry mainly comes from the value of animal husbandry. Development and utilization are conducive to increasing the disposable income of people in the region. In terms of social benefits. The four indicators of grassland area(S1), grassland availability factor(S2), population growth rate(S3), and number of people out of poverty(S4) are taken into consideration. We know that Golog Zang A.P belongs to the San jiang yuan Natural Ecological Reserve. The area contains a large number of grassland resources. Grazing is the main source of income for residents. However, in recent years, with the excessive use of grassland and the irrational use of water resources, it has seriously affected the sustainable development of the region's economy. With the increase of the population, these problems will have a greater negative effect on regional stability if they are not reasonably solved. From the indicator of the number of people who are out of poverty, we can clearly see the social and economic development of the region. We cannot judge the size of regional water security problems by the amount of regional water resources. Analyzing and analyzing the benefits of water resources to our economy and society can help us formulate targeted policies to solve a series of problems that arise from this.

The raw data is shown in the following table. A, B, C, D, E, and F in the table represent Maqin County, Banma County, Gande County, Dari County, Jiuzhi County, and Maduo County, respectively.

**Table 2.** Indicator Data.

Area	E1 (yuan)	E2 (10 <sup>4</sup> yuan)	E3 (10 <sup>4</sup> yuan)	S1 (10 <sup>4</sup> hectares)	S2 (%)	S3 (‰)	S4 (person)
A	9834	21885.8	165.34	117.58	9	11.52	1025
B	11843	13107.33	157.72	39.67	85	7.15	2413
C	9783	13242.18	203.65	65.12	90	11.7	5715
D	12266	12175.46	194.27	140.166	79	13.57	4082
E	9745.69	12853.42	158.48	73.236	90	12.84	2262
F	15602	8295.29	226.86	229.91	79	13	2698

### 3.1.3. Evaluation Methods

This evaluation uses fuzzy comprehensive evaluation as the evaluation model of social and economic benefits, and uses entropy weight method to calculate the index weight. According to the mathematical principle of fuzzy relation, fuzzy information is quantified through fuzzy information, which can effectively quantify the influencing factors with unclear boundaries and concepts.

#### 3.1.3.1. Entropy Weight Method to Calculate Index Weight [11]

##### (1) Standardized treatment of evaluation indicators

Among the selected evaluation indicators. There are indicators that have a positive impact on the efficient use of water resources, that is, positive indicators; There are also indicators that have a negative impact on the efficient use of water resources, that is, negative indicators. Therefore, the initial data needs to be standardized for comparability.

For positive indicators, the standardized processing formula is:

$$X_{ij} = \frac{X_{ij} - \min(X_{ij})}{\max(X_{ij}) - \min(X_{ij})} \quad (1)$$

For negative indicators, the standardized processing formula is:

$$X_{ij} = \frac{\max(X_{ij}) - X_{ij}}{\max(X_{ij}) - \min(X_{ij})} \quad (2)$$

In the formula: X is the data in Table 2, i and j are respectively represented as the i-th row and the j-th column.

##### (2) Determination of entropy

The entropy value is calculated by the following formula:

$$H_i = -k \sum_{j=1}^n f_{ij} \ln f_{ij} \quad (3)$$

$$f_{ij} = \frac{g_{ij}}{\sum_{j=1}^n g_{ij}} \quad (4)$$

$$k = \frac{1}{\ln n} \quad (5)$$

In the formula:  $H_i$  is the entropy value of evaluation index  $I$ , and it is assumed here that when  $f_{ij}=0$ ,  $f_{ij}\ln f_{ij} = 0$ ;  $g_{ij}$  is the normalized matrix of the original evaluation index,  $K$  is Boltzmann constant, and  $n$  is the total number of research objects. The smaller the entropy value, the more useful information the index can provide for decision makers.

### (3) entropy weight of evaluation index

After obtaining the entropy value  $H$ , the entropy weight of the evaluation index is determined by the following formula:

$$\omega_i = \frac{1-H_i}{m - \sum_{i=1}^m H_i} \quad (0 \leq \omega_i \leq 1, \sum_{i=1}^m \omega_i = 1) \quad (6)$$

#### 3.1.3.2. Steps to Establish Fuzzy Mathematical Model

##### (1) establish fuzzy evaluation matrix

According to Zadeh's groundbreaking research in 1965[12], he proposed the concept of fuzzy sets and Chen Shouyu's fuzzy set theory[13,14]. It is assumed that the system has  $n$  candidate options to form the solution set, and  $m$  indicators to perform the solution set. The set of indicators to be judged, then the eigenvalue matrix of the indicator is

$$X_{m \times n} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} = (x_{ij})_{m \times n} \quad (7)$$

The fuzzy evaluation matrix can be obtained by evaluating the indexes to determine the membership degree of the evaluation objects to the elements of the evaluation set [15].

$$R_i = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \quad (8)$$

$(n = 5; m = \text{number of indicator})$

##### (2) establishing fuzzy comprehensive evaluation

The fuzzy comprehensive evaluation of the corresponding indicators can be obtained through the weight set  $\omega_i$  and the single-factor evaluation matrix  $R_i$  of each indicator.

$$B_i = \omega_i \times R_i, \quad i = 1, 2 \quad (9)$$

The above evaluation results are regarded as a fuzzy comprehensive evaluation matrix  $B$  composed of multiple single-factor evaluation sets, and the evaluation of the benefit is as follows:

$$C = A \times B \quad (10)$$

##### (3) Evaluation of results

It can be known from the above calculation results that the comprehensive evaluation vector of water resources is converted into the price of water resources by the following formula, thereby reflecting the value of water resources.

$$W = C \times S^T \quad (11)$$

Where  $W$  is the price of water resources and  $S^T$  is the transposed vector of water resource price vectors,  $S = (P \ P_1 \ P_2 \ P_3 \ 0)$ .

The determination of the vector  $S$  uses the method of social affordability. The affordability mentioned in this paper refers specifically to the tolerance of people to accept this fact under the stimulus of rising water prices. In the vector  $S$ ,  $P$  is the upper limit of the water price, that is, the highest tolerable water price. The calculation formula of  $P$  is

$$P = BE/C-D \quad (12)$$

In the above formula,  $B$  is the maximum water fee bearing index (the calculation formula is  $B = \text{water fee expenditure/actual income}$ ),  $E$  is the actual income of residents,  $C$  is the water consumption of residents, and  $D$  is the cost of water supply and normal profit for tap water.

## 4. Result Analysis

### 4.1. Analysis of Basic Data

By analyzing the data in Table 2, we can clearly see that Maduo county is in the leading position in Chinese and Tibetan medicinal materials, per capita disposable income and grassland area. But the output value of agriculture and animal husbandry is the lowest. Because the county actively adjusts its industrial structure, the proportion of tertiary industry investment in fixed assets investment reaches 80.2%. Maqin County is at the highest value in agricultural and animal husbandry output value. Gande county and jiuzhi county are both at a high level in grassland availability coefficient. In terms of population growth rate, Dari County has the highest rate, while Gande county has the highest number of people out of poverty. However, comparing various data, it can be clearly found that there is not much difference between counties, which increases the reliability of the evaluation of social and economic benefits in the region. Through per capita disposable income, it can be seen that the region's economy is still backward and the people's living standard needs to be improved.

### 4.2. Analysis of Weight Calculation Results

We process the data in Table 2 and use the above calculation steps to calculate by programming, we can get the following calculation results. The weight calculated by entropy weight method are as follows:

**Table 3.** Weight calculation results.

Name	weight
Output Value of Chinese and Tibetan Medicinal Materials	0.347
Output value of agriculture and animal husbandry	0.196
Per capita disposable income	0.457
Grassland area	0.372
Grassland availability factor	0.166
population rate of increase	0.174
Number of people out of poverty	0.288

In economic benefits, from Table 3, agriculture and animal husbandry have the lowest weight, that is, this index has the greatest reference significance, which coincides with the actual situation that agriculture and animal husbandry are the main economic sources of

local herdsmen in high altitude and low temperature areas, thus increasing the reliability of benefit evaluation. In social benefits, there is little difference in the weight of various indicators. The grassland availability coefficient and population growth rate are relatively low. Among them, the grassland availability coefficient directly affects the grazing area of herdsmen. Population growth rate, as the main reference data to measure the population development in a region, is more valuable for poor regions, reasonably controls the population and avoids vicious circle.

#### 4.3. Fuzzy Comprehensive Evaluation Results

The following results can be calculated by Equation 5-8:

Single factor evaluation matrix of economic benefits

$$R^1 = \begin{bmatrix} 0.2 & 0.4 & 0.3 & 0.1 & 0 \\ 0.4 & 0.3 & 0.1 & 0.1 & 0.1 \\ 0.3 & 0.3 & 0.2 & 0 & 0.2 \end{bmatrix}$$

Single factor evaluation matrix of social benefits

$$R^2 = \begin{bmatrix} 0.2 & 0.3 & 0.2 & 0.1 & 0.2 \\ 0.3 & 0.2 & 0.2 & 0.2 & 0.1 \\ 0 & 0.3 & 0.3 & 0.3 & 0.1 \\ 0.4 & 0.2 & 0.1 & 0 & 0.3 \end{bmatrix}$$

Through the weight set of each index and the single factor evaluation matrix RI, the fuzzy comprehensive evaluation of the corresponding index can be obtained:

$$B_1 = \omega_1 \times R^1 = (0.2849, 0.3347, 0.2151, 0.0543, 0.111)$$

$$B_2 = \omega_2 \times R^2 = (0.2394, 0.2546, 0.1886, 0.1226, 0.1948)$$

Considering the above evaluation results as a fuzzy comprehensive evaluation matrix B composed of multiple single-factor evaluation sets, the economic and social benefit evaluation is:  $C = A \times B = (0.252, 0.276, 0.196, 0.104, 0.172)$ .

#### 4.4. Value of Water Resources in Terms of Residents' Lives

In 2018, the per capita disposable income of Golog Zang A.P was 14557 yuan, the per capita water consumption was about 66 L per day, and the cost and profit of tap water supply was 2.1 CNY/m<sup>3</sup>, maximum water fee bearing index is calculated as 0.03.

Obtainable:  $P = (0.03 \times 14557) \div (0.066 \times 365) - 2 = 16.13$  CNY/m<sup>3</sup>, Equal interval method, calculated price vector  $S = [16.13 \ 12.0975 \ 8.065 \ 4.0325 \ 0]$ .

Through the calculation results, we can see that under the comprehensive influence of various factors, the price of domestic water in Golog Zang A.P is around 9.4 CNY/m<sup>3</sup>.

## 5. Economic Value of Water Resources in Other Industries

Calculate the value of water resources for irrigation crops, industry, construction and the tertiary industry by using the theory of emergy and the benefit sharing coefficient, the regional water resources' contribution to the economy can be obtained more accurately.

### 5.1. System Emergy Calculation

Emergy analysis theory and method was founded in the 1980s by H.T. Odum, a famous American ecologist and pioneer of system energy analysis. Emergy is a new scientific concept and metric, defined as "the quantity of another form of energy contained in a kind of flowing or stored energy", or "the quantity of an efficiency that is directly and indirectly applied in the formation of products and services". Its essence is to include energy or embodied energy. Collect and organize the raw data of the research object about energy flow, material flow and money flow, distinguish and classify them, and calculate their energy.

Emergy calculation formula[16]:

$$M = \tau \times B \quad (13)$$

In the formula: M represents the energy value (sej);  $\tau$  represents the energy value conversion rate (sej/J or sej/g); B represents the energy or mass of the material (J or g).

Formula for calculating benefit sharing coefficient:

$$\varepsilon = \frac{M_i}{\sum_{j=1}^n M_j} \quad (14)$$

In the above formula:  $M_i$  is the total energy value of the water supply;  $M_j$  is the energy value of various inputs in the system;  $\varepsilon$  is the benefit sharing factor.

### 5.2. Irrigation Planting

The formula for calculating the economic value of water resources for irrigation planting is[17], Precise data are shown in Table 4:

$$EVW_{I_r} = TVW_{I_r} / Q_{I_r} \quad (15)$$

$$TVW_{I_r} = \varepsilon_{I_r} \times \omega \times Y \times P \quad (16)$$

In the formula:  $EVW_{I_r}$  is the economic value of irrigation water resources (CNY/m<sup>3</sup>);  $TVW_{I_r}$  is the irrigation sharing benefit of the planting industry ( $10^8$ CNY);  $Q_{I_r}$  is the irrigation water volume ( $10^8$ m<sup>3</sup>);  $\varepsilon_{I_r}$  is the apportionment coefficient of crop irrigation benefits;  $\omega$  is the planting area of crops (hm<sup>2</sup>); Y is the average yield per year of crops with irrigation engineering (kg/hm<sup>2</sup>); P is the product price of crops (CNY/kg).

**Table 4.** Emergy analysis table of crop production system in Golog Zang A.P.

Project	Emergy conversion rate(sej/unit)
Solar energy	1
Wind energy	623
Irrigation water	$1.74 \times 10^{12}$
Agricultural machinery	$7.5 \times 10^7$
Diesel oil	$6.6 \times 10^4$
Nitrogen	$4.62 \times 10^9$

Phosphate fertilizer	$1.78 \times 10^{10}$
Potassium fertilizer	$1.74 \times 10^9$
Compound fertilizer	$2.80 \times 10^9$
Pesticides	$1.62 \times 10^9$
Agricultural film	$3.8 \times 10^8$

Note: (1) The original data are from the statistical yearbook. (2) The energy conversion rate and calculation formula are from the reference [18,19].

**Table 5.** Unilateral water value of irrigation planting in the research area.

Years	Coefficient	Unilateral water value(yuan/m <sup>3</sup> )
2009	0.51	18.77
2010	0.485	15.24
2011	0.486	20.46
2012	0.45	22.06
2013	0.43	25.73
2014	0.41	28.04
2015	0.38	28.99
2016	0.44	33.15
2017	0.39	40.89
2018	0.38	39.28

From Table 5, we can see that the value of local irrigation water is relatively high in recent years, reaching about 40 CNY /m<sup>3</sup>. The reason is that as the awareness of water resources increases and the application of drip irrigation technology changes the backward methods such as flood irrigation and string irrigation in the past. Avoiding water and soil loss and large amount of water evaporation, improving the efficiency of water resources utilization, and thus increasing the value of water resources.

### 5.3. Industry, Construction and the Tertiary Industry

The formula for calculating the economic value of water resources for industry, construction and the tertiary industry is [19]. Precise data are shown in Table 6:

$$EVW_S = TVW_S / Q_S \quad (17)$$

$$TVW_S = B_S \times \varepsilon_{W,S} + F_{W,S} \quad (18)$$

In the formula:  $EVW_S$  represents the economic value of water resources of industry S (yuan/m<sup>3</sup>);  $TVW_S$  is the shared benefit of water withdrawal by industry S ( $10^8$ yuan);  $Q_S$  is the water withdrawal of industry S ( $10^8$ m<sup>3</sup>);  $B_S$  is the total profit and tax of industry S( $10^4$ yuan);  $\varepsilon_{W,S}$  is the benefit allocation coefficient of water withdrawal by industry S;  $F_{W,S}$  is the complete water supply cost of industry( $10^4$ yuan)

**Table 6.** Emergy analysis tables for production systems in other industries.

Project	Emergy conversion rate(sej/unit)
Solar energy	1
Wind energy	623
Labor cost	$7.44 \times 10^{11}$
Energy (equivalent to standard coal)	39800
Tap water	$3.89 \times 10^{13}$
Surface water	$1.63 \times 10^{12}$
Groundwater	$1.85 \times 10^{12}$
Investment in fixed assets	$7.44 \times 10^{11}$

Note: (1) The original data are from the statistical yearbook. (2) The energy conversion rate and calculation formula are from the reference[20].

**Table 7.** Corresponding value of unilateral water in other industries in the study area.

years	Industrial efficiency coefficient	Unilateral water value (yuan/m <sup>3</sup> )	Construction industry efficiency coefficient	Unilateral water value (yuan/m <sup>3</sup> )	Tertiary Industry efficiency Coefficient	Unilateral water value (yuan/m <sup>3</sup> )
2014	0.064	24.58	0.04	26.89	0.078	21.66
2015	0.07	15.39	0.02	13.96	0.073	36.84
2016	0.068	22.12	0.027	25.11	0.074	39.02
2017	0.067	26.86	0.026	25.27	0.067	40.23
2018	0.06	25.4	0.038	24.32	0.067	41.76

From the table 7 we can clearly see that the corresponding value of industrial unilateral water is between 15.39~26.86 CNY/m<sup>3</sup>, and the average value is around 22.9 CNY/m<sup>3</sup>. The value of the construction industry is around 23.11 CNY/m<sup>3</sup>. The value of the tertiary industry is around 40 CNY/m<sup>3</sup>. We can see that the unilateral water efficiency corresponding to the service industry is the highest, which helps us to make reasonable industrial and water resource adjustments. Furthermore, improve people's living standards finally.

## 6. Discussion

According to the actual situation of the study area, this paper collects economic and social data in the study area, and uses fuzzy mathematics and energy theory to calculate the value of its water resources. The entropy weight method is used to calculate the index weight, and the value of unilateral domestic water is calculated by using fuzzy mathematics method at about 9.4 CNY/m<sup>3</sup>. Based on energy theory, the benefit sharing coefficients of other industries are calculated, and the unilateral water value of irrigation, industry, construction, and tertiary industry is around 37 CNY /m<sup>3</sup>, 22.9 CNY /m<sup>3</sup>, 23.11 CNY /m<sup>3</sup>, and 40 CNY/m<sup>3</sup>.

(1) There is a big gap between the calculation results of water resources value and the actual water price in the area. At this stage, the residents' water fee tolerance index is only 0.8%, which is still a long way from the standard 3%, Therefore, raising water prices will not have a big impact on the normal life of residents.

(2) There is a big gap between the water price and actual value of industrial and manufacturing industries and the tertiary industry, and the low water efficiency has caused the waste of regional water resources to a certain extent.

## 7. Conclusions and Future Research

The pricing method of water resources based on the traditional natural resource value theory cannot objectively and comprehensively interpret the value of water resources. The

emergy analysis theory divides the energy flow, material flow and information that are not unified but closely related in the ecological economic system. The flow and currency flow are quantitatively analyzed using the common measure of emergy. This article uses emergy theory and fuzzy mathematics to quantitatively evaluate the value of regional water resources, discusses basic concepts and research methods, and proposes an emergy analysis and calculation model for regional water resources value evaluation, and combines the water resources of the Golog Zang A.P a case study was conducted.

The research results are consistent with the market value theory, indicating that the regional water resources value evaluation method based on emergy theory and fuzzy mathematics has a certain feasibility and scientific rationality, and can be used for the formulation and adjustment of regional water prices and the sustainable utilization of water resources. And optimized configuration provides a scientific theoretical reference.

However, considering the errors and incompleteness of the relevant data used in this research due to certain differences in survey and statistical methods, The calculation can only be simplified with the fixed value of the existing research results, coupled with the complexity of water resource value research and the application of emergy theory in resource value research is not yet mature. Therefore, the application of emergy theory and fuzzy mathematics to the quantitative evaluation of the value of regional water resources is only an attempt to study the value of water resources, and further research and improvement are needed.

## References

- [1] Zhu JL, Tao XY, Wang SJ, Wu JP. Measurement and Analysis of Water Resources Value in the Huaihe River Basin. *Journal of Natural Resources*. 2005; 20(1):126-131.
- [2] He J, Chen XK. Calculation of shadow price of dynamic water resources in 9 river basins in China. *Water Resources Economy*. 2005; 23(1):14-19.
- [3] Jiang WL. *Water Resources Value Theory*. Science Press, Beijing, 1998.
- [4] Miao HY, Yang ZJ. Fuzzy comprehensive evaluation of regional water resources value. *South-to-North Water Transfer and Water Conservancy Science and Technology*. 2003; 1(15):17-19.
- [5] Robert CJ. Micro and macro-level approaches for assessing the value of irrigation water. *World Bank Policy Research Working*, 2005, 3778.
- [6] Glenn-Marie L, Rashid H. The economics of water management in southern Africa: An environmental accounting approach. Edward Elgar Publishing Limited, 2006.
- [7] Pan X. Social and ecological accounting matrix: an empirical study for China. *Int. Conf. IaO technique*, 2009.
- [8] Thorbeck. The use of social and ecological accounting matrices in modelling. 26th Gen. Conf. Int. Assoc. for Research in Income and Wealth. Cracow, Poland, 2002.
- [9] Krajnc D, Glavic P. A model for integrated assessment of sustainable development. *Resources Conservation & Recycling*. 2005; 43(2):189-208.
- [10] Ni JP, Ping L, Wei CF, Xie DT. Evaluation of Regional Land Development and Consolidation Potential Based on AHP and Entropy Weight Method. *Journal of Agricultural Engineering*. 2009; 25(05):202-209.
- [11] Lu TC, Kang K. Application of Entropy Method and Analytic Hierarchy Process in Weight Determination. *Computer Programming Skills and Maintenance*. 2009; 22:19-20+53.
- [12] Zadeh LA. Fuzzy sets. *Information and Control*. 1965; 8(3): 338-353.
- [13] Wang WC, Xu DM, Chen SY, Qiu L. Research progress of variable fuzzy set theory and its application in water science. *Advances in Water Conservancy and Hydropower Science and Technology*. 2012; 32(05):89-94.
- [14] Chen SY. Variable Set and Variable Set Principle and Method for Optimal Selection Decision of Water Resources System. *Journal of Water Resources*. 2012; 43(09):1066-1074.
- [15] Zhang LW, Ding HF, Chen YJ. Railway Dangerous Goods Transportation Early Warning Based on Entropy Weight Fuzzy Comprehensive Evaluation. *China Safety Science Journal*. 2012; 22(05):119-125.

- [16] Ni HZ, Zhang CL, Chen GF, Zhao J, Feng L. Problems and Reform Suggestions of Water Supply Costs and Prices in China's Cities. *Water Resources and Hydropower Engineering*. 2019; 50(8):209-215.
- [17] Tang JR. *Economic Economics*. Chemical Industry Press, Beijing, 2005.
- [18] Odum HT. *Environmental Accounting: Emergy and Environmental Decision Making*. John Wiley & Sons, New York, 1996.
- [19] Buenfil AA. *Emergy Evaluation of Water*. Department of Environmental Engineering Sciences, University of Florida, Gainesville, U. S, 2001.
- [20] Li LX, Gan H, Wang L, Ni HZ, Mu SM. Calculation and Analysis of Economic Value of Water Resources. *Journal of Natural Resources*. 2008; 3: 494-499.

# Integrated Flood Risk Assessment System Based on the Set Pair Analysis and Variable Fuzzy Sets for Cascade Dams

Yu CHEN<sup>a,b,1</sup>

<sup>a</sup>*State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Chengdu 610065, China*

<sup>b</sup>*Institute for Risk and Disaster Reduction, University College London, Gower Street, London WC1E 6BT, United Kingdom*

**Abstract.** Flood-caused dam break is an important dam safety problem. Dam flood risk assessment based on an appropriate assessing method is hence essential for reducing dam-break flood damages and losses. In this research, a comprehensive flood risk assessment system for cascade dams was developed by combining the set pair analysis (SPA) and variable fuzzy sets (VFS) theory (namely SPA-VFS for short), and expressing the flood risk as a product of hazard and vulnerability, which were quantified by selected indicators. In the case study of the Dadu river basin, reservoir capacity, dam height, dam age, gross domestic product (GDP) density, population density, and predicted damage were selected as the evaluation indicators, and the evaluation results derived the flood risk classes of each studied dam. The study indicates that the proposed SPA-VFS-based flood risk assessment model is simple and feasible, which enables its further application in flood risk analysis and evaluation, and its extension in the risk assessment of other natural disasters.

**Keywords.** Flood, set pair analysis (SPA), variable fuzzy sets (VFS), risk assessment, cascade dams

## 1. Introduction

Flood is one in every of the foremost cosmopolitan natural disasters to life and property in the world. The dam break caused by flood often brings catastrophic damages and enormous impacts to humans, society, economy, and environment, and becomes a worldwide problem. The gradually increasing flood occurrence provokes the development of dam safety management and flood risk decision-making, which is closely based on high-accuracy dam flood risk assessment.

Many approaches, such as uncertainty analysis, statistical analysis, and multi-criteria decision-making analysis, have been applied in risk assessment [1]. In recent years, more and more researchers pay attention to the uncertainty method, such as the variable fuzzy sets (VFS) theory and set pair analysis (SPA) method. In this context, an integrated flood

---

<sup>1</sup> Corresponding Author: Yu Chen, State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Chengdu 610065, China; E-mail: [rainchen393@hotmail.com](mailto:rainchen393@hotmail.com)

This work was supported by Sichuan Science and Technology Program [grant number 2019YFH0140]; International Visiting Program for Excellent Young Scholars of SCU; and the National Natural Science Foundation of China [grant numbers 41501554].

risk system combining SPA and VFS methods was constructed and implemented for cascade dams in the study area.

## 2. Methodology

### 2.1. Variable Fuzzy Sets (VFS)

The VFS theory based on the relative difference function is a convenient tool for processing random, fuzzy, and multi-indicator problems, and has been used in the field of risk evaluation. Denote the amount of sample sets by  $n$ , then

$$X = \{x_1, x_2, \dots, x_n\} \quad (1)$$

where  $x$  is the sample set.

Let there are  $m$  indicators, the indicator eigenvalue of  $j$ th sample is represented as

$$x_j = \{x_{1j}, x_{2j}, \dots, x_{mj}\} \quad (2)$$

Based on  $i$ th ( $i=1, 2, \dots, m$ ) indicator eigenvalue of  $j$ th ( $j=1, 2, \dots, n$ ) sample,  $x_{ij}$ , , the sample set can be defined as follows.

$$X = (x_{ij}) \quad (3)$$

To define VFS, let  $u$  is the arbitrary element and  $U$  is a fuzzy concept set, for  $\forall u \in U$  ,  $\mu_A(u), 0 \leq \mu_A(u) \leq 1$  and  $\mu_{A^c}(u), 0 \leq \mu_{A^c}(u) \leq 1$  are relative membership degrees of  $u$  to the attractive intervals  $A$  and  $A^c$ , respectively, describing attractability and repellency respectively, and  $\mu_A(u)+\mu_{A^c}(u)=1$  . Define  $V$  as VFS,  $V = \{(u, D_A(u)) | u \in U, D_A(u) = \mu_A(u) - \mu_{A^c}(u), D_A(u) \in [-1, 1]\}$  as VFS, where  $D_A(u)$  is the relative difference degree of  $u$  to  $A$ . Furthermore, suppose  $A_+$ ,  $A_-$ , and  $A_0$  are attracting, repelling and qualitative change sets of  $V$ , respectively [2], then,  $A_+ = \{u | u \in U, D_A(u) \in (0, 1)\}$  ,  $A_- = \{u | u \in U, D_A(u) \in (-1, 0)\}$  , and  $A_0 = \{u | u \in U, D_A(u) = 0\}$  .

Considering the advantage of VFS for expressing the nonlinear relationship between evaluation indicators and risk grades, and the limitation of VFS for solving assessment indicators induced information duplication, more and more theories and methods are used to combine with VFS for various improvements.

### 2.2. Set Pair Analysis (SPA)

SPA is a new theory of expressing and tackling the systematic uncertainty by integrating uncertainty and certainty as to the certain-uncertain system, dividing the certain-uncertain relationship into identity, contrary and discrepant, and utilizing a connection number to represent degrees of identity, contrary and discrepancy in specific cases based on analyzing characteristics of the studied set pair. SPA is used in this paper to calculate the relative membership degree. Let  $H(A, B)$  is a set pair formed by putting together two

sets  $A$  and  $B$ , the connection degree  $\mu_{(A-B)}$  of the set pair can be expressed as the following three-element connection form [3-4]:

$$\mu_{(A-B)} = a + bi + cj \quad (4)$$

The general form of connection degree shown in Eq. (4) can be extended to the five-element connection degree as follows [5]:

$$\mu_{(A-B)} = a + bi + cj = a + b_1 i_1 + b_2 i_2 + b_3 i_3 + cj \quad (5)$$

where  $a + b_1 + b_2 + b_3 + c = 1$ ,  $a; b_1, b_2, b_3; c$  are components of identity degree; discrepancy degree; and contrary degree, respectively, and used to determine the grades of evaluation indicators in flood risk assessment;  $i_1, i_2, i_3$  are uncertainty coefficients of discrepancy degree and  $i_1, i_2, i_3 \in [-1, 1]; j = -1$ .

### 2.3. Flood Risk Assessment Based on SPA-VFS

The VFS based on the membership function is limited by complicated calculation and multi-variables requirements for deriving the discrepancy degree. SPA provides a new and simple way to establish the connection degree, which is basically the same as the discrepancy degree. The two methods hence can be combined to form a SPA-VFS flood risk assessment system, which can be implemented in the following steps [6]:

(1) Determine assessment indicators and standards of the dam flood risk. Denote the assessment sample as  $\{x_j | j = 1, 2, \dots, m\}$  and corresponding evaluation grade standard as  $\{s_{jk} | j = 1, 2, \dots, m; k = 1, 2, \dots, n\}$ , where  $m, n$  are the total numbers of indicators and grades, respectively.  $x_j$  is the eigenvalue of the indicator  $j$ ,  $s_{jk}$  is the evaluation standard for the indicator  $j$  and level  $k$ .

(2) Calculate the indicator weights  $\{w_j | j = 1, 2, \dots, m, \sum_{j=1}^m w_j = 1\}$  by using the analytic hierarchy process (AHP) method, where  $w_j \in [0, 1]$  is the weight of the indicator  $j$ . More details about AHP can reference the research of Orencio and Fujii [7].

(3) Derive the connection degree between two sets of evaluation indicators and assessment levels by using SPA. If the evaluation indicators of samples  $l_1, l_1+1, \dots, l_1+l$  are in the grade  $k$ , then the identity degree  $a$  is defined as  $a = \sum_{j=l_1}^{l_1+l} w_j$ . If the evaluation indicators of samples  $l_1+l+1, l_1+l+2, \dots, l_1+l+l$  are located in the grade  $k$  interval, the contrary degree  $c$  is denoted as  $c = \sum_{j=l_1+l+1}^{l_1+l+l} w_j$ . If the evaluation indicators of samples  $l_1+l+1, l_1+l+2, \dots, l_1+l+m$  are located in the grade  $k$  adjacent, the discrepancy degree  $b_j$  of a single indicator is expressed as  $b_j = w_j$ , ( $j = l_1+l+1, l_1+l+2, \dots, l_1+l+m$ ). And the difference degree coefficient  $I_j$  for a single indicator is

$$I_j = \begin{cases} 1 - 2 \cdot \left| \frac{s_{j(k-1)} - x_j}{s_{j(k-1)} - s_{j(k-2)}} \right| & x_j \in \text{grade}(k-1) \\ 1 - 2 \cdot \left| \frac{x_j - s_{jk}}{s_{j(k+1)} - s_{jk}} \right| & x_j \in \text{grade}(k+1) \end{cases} \quad (6)$$

where  $x_j \in \text{grade}(k-1)$  and  $x_j \in \text{grade}(k+1)$  represent the indicator values locate in the grade  $k-1$  adjacent and grade  $k+1$  adjacent, respectively.

The connection degree  $u_j$  between  $x_j$  and the grades  $k$  is

$$u_j = a + \sum_{j=l_2+1}^m b_j \cdot I_j + c \cdot J \quad (7)$$

where  $J=-1$ .

Then the synthetic connection degree  $u$  can be calculated as [4]:

$$u = \sum_{j=1}^m w_j \cdot u_j \quad (8)$$

where  $u \in [-1, 1]$ .

(4) Calculate the relative membership degree  $v_k$  based on the relation degree  $u_k$  of grade  $k$ .

$$v_k = 0.5 + 0.5 \cdot u_k \quad (9)$$

(5) Calculate the assessment grade feature value  $H$  by using the rank feature value method [8]:

$$H = \sum_{k=1}^n k \cdot (v_k / \sum_{k=1}^n v_k) \quad (10)$$

In this study, flood risk grades are illustrated as lowest, low, moderate, high, and highest, denoted by 1, 2, 3, 4, and 5, respectively. Based on the calculation results of Eq. (10), the qualitative grades can be assigned to each dam according to the following rule [9]:

$$\text{risk grade} = \begin{cases} 1, & 1.0 \leq H \leq 1.5, \\ 2, & 1.5 < H \leq 2.5, \\ 3, & 2.5 < H \leq 3.5, \\ 4, & 3.5 < H \leq 4.5, \\ 5, & 4.5 < H \leq 5.0, \end{cases} \quad (11)$$

### 3. Case study

The Dadu river basin is located in the southwest of China, where one of the thirteen hydropower bases is distributed. Currently, there are a total of 22 planned, under-constructed, and built cascade reservoirs in the mainstream, wherein Shuangjiangkou and Pubugou dams are the control hydraulic projects of the cascade reservoir system, and situated in the upstream and the downstream of the basin, respectively. From upstream to downstream, Busigou (BSG), Shuangjiangkou (SJK), Luding (LD),

Dagangshan (DGS), Longtoushi (LTS), and Pubugou (PBG) dams are selected as study examples, and each dam is regarded as a basic assessment unit.

According to the disaster risk theory, the flood risk can be expressed as a product of flood hazard and flood vulnerability. Correspondingly, the studied flood risk assessment indicators should be selected from the aspects of disaster-inducing factors, disaster-forming environment, and disaster-bearing bodies. Considering engineering properties of cascade dams and the natural quality of floods, six indicators are chosen, including three hazard indicators of reservoir capacity ( $C_1$ ), dam height ( $C_2$ ) and dam age ( $C_3$ ), and three vulnerability indicators of GDP density ( $C_4$ ), population density ( $C_5$ ), and predicted damage ( $C_6$ ). It should be noted that  $C_6$  is a qualitative indicator, and should be characterized by the highest water level of the reservoir (HWL).

To quantify these risk indicators, the historical flood information, statistical yearbook, field survey data, and expert judgments are collected. After that, each indicator is divided into five levels from lowest (1) to highest (5), and the division is shown in Table 1.

The indicator weights are calculated by the AHP method as:

$$w_1 = (0.190, 0.111, 0.032, 0.059, 0.130, 0.478) \quad (12)$$

It can be seen from the weighting result that the predicted damage ( $C_6$ ) indicator with the highest weight is the most important risk influencing factor, while the dam age ( $C_3$ ) indicator is the least important one.

After obtaining the indicator weights, Eqs. (6)-(10) was applied step by step to calculate the single and synthetic connection degrees, the relative and normalized membership degrees, and the assessment grade feature values. The final risk grade of each dam was determined in the light of Eq. (11) and shown in Table 2.

The results show that dams located in the upper reaches possess low risk, while dams distributed in the middle reaches possess moderate risk. Low-risk dam locations are relatively low-developed economic and sparse population distribution areas. For midstream areas with high GDP and population density and an intense rainfall zone, where the safety performance of the dams is worse than that of upstream dams, which is reflected by the relatively higher grade feature values and risk grade ratings. Specifically, the dam height of the DGS dam is more than 200 m, and the reservoir capacity of the PBG dam is more than 5 billion  $m^3$ . The SJK dam is a control project of the basin cascade system, plays the role of energy production and flood prevention, and is the most sensitive cascade because its failure will induce catastrophic damage for the downstream cascade reservoirs.

**Table 1.** Level standards of flood risk assessment indicators.

Risk indicators	Lowest	Low	Moderate	High	Highest
Reservoir capacity ( $10^6 m^3$ )	<1	1-10	10-100	100-1000	>1000
Dam height (m)	<15	15-70	70-166	166-300	>300
Dam age (years)	<5	5-45	45-85	85-115	>115
GDP density (billion yuan/ $km^2$ )	<10	10-40	40-80	80-100	>100

Population density (population/km <sup>2</sup> )	<10	10-100	100-250	250-500	>500
Predicted damage (highest water level of reservoir (m))	<UWL <sup>a</sup>	UWL-DFL <sup>b</sup>	DFL-CFL	CFL-DH <sup>c</sup>	>DH
<sup>a</sup> UWL=Upper water level for flood control;					
<sup>b</sup> DFL=Design flood level;					
<sup>c</sup> CFL=Checked flood level; DH=Dam height.					

**Table 2.** Results of the flood risk grades of the studied dams.

Normalized membership	BSG	SJK	LD	DGS	LTS	PBG
1	0.190	0.521	0.239	0.062	0.127	0.058
2	0.438	0.156	0.318	0.188	0.246	0.159
3	0.334	0.140	0.112	0.304	0.381	0.427
4	0.038	0.100	0.011	0.328	0.246	0.341
5	0.000	0.083	0.320	0.118	0.000	0.015
Grade feature value	2.2	2.068	2.855	3.252	2.746	3.096
Risk grades	Low	Low	Moderate	Moderate	Moderate	Moderate

#### 4. Conclusions

This study takes cascade dams, BSG, SJK, LD, DGS, LTS, and PBG, in the Dadu river basin as an example to conduct an integrated flood risk assessment. According to the disaster risk theory and actual situation of the study case, three hazard indicators and three vulnerability indicators are selected and then weighted by the AHP method. The integrated flood risk assessment of each studied dam was implemented by the built SPA-VFS model, and the resulting flood risk grades of each dam indicated that the cascade dams are safe. The SPA-VFS retains the advantages of SPA and VFS methods, can simplify the calculation of the discrepancy degree using SPA and identify the membership relationship between assessment indicators and standards utilizing VFS theory, and overcome the limitations of the maximum membership degree principle (MMDP) in traditional SPA method.

#### References

- [1] Jiang WG, Deng L, Chen LY, Wu JJ and Li J. Risk assessment and validation of flood disaster based on fuzzy mathematics. Progress in Natural Science. 2009; 19:1419-1425.
- [2] Chen SY, Guo Y. Variable fuzzy sets and its application in comprehensive risk evaluation for flood-control engineering system. Fuzzy Optim Decis Mak. 2006; 5:153-162.
- [3] Zou Q, Zhou JZ, Zhou C, Song LX and Guo J. Comprehensive flood risk assessment based on set pair analysis-variable fuzzy sets model and fuzzy AHP. Stochastic Environmental Research Risk Assessment. 2013; 27: 525-546.
- [4] Guo EL, Zhang JQ, Ren XH, Zhang Q and Sun ZY. Integrated risk assessment of flood disaster based on improved set pair analysis and the variable fuzzy set theory in central Liaoning Province, China. Natural

- Hazards. 2014; 74:947-965.
- [5] Chen Y and Shu LC. Set pair analysis for regional land subsidence vulnerability assessment. 2012 International Conf. on Computer and Information Science, Safety Engineering (CAISSE 2012), June 15-17, 2012, Wuhan.
- [6] Zhou JF. SPA-fuzzy method based real-time risk assessment for major hazard installations storing flammable gas. Safety Science. 2010; 48:819-822.
- [7] Orencio PM and Fujii M. A localized disaster-resilience index to assess coastal communities based on an analytic hierarchy process (AHP). International Journal of Disaster Risk Reduction. 2012; 3:62-75.
- [8] Chen SY and Guo Y. Variable fuzzy sets and its application in comprehensive risk evaluation for flood-control engineering system. Fuzzy Optimization and Decision Making. 2006; 5:153-162.
- [9] Guo Y 2006 Research on application of engineering fuzzy sets theory in water resources and flood control system. Dalian University of Technology, Dalian, 2006.

This page intentionally left blank

## Part 2.

# Machine Learning and Intelligent Systems III

This page intentionally left blank

# Causal Reasoning Model Based on Medical Knowledge Graph for Disease Diagnosis

Ze XU <sup>a</sup>, Huazhen WANG <sup>a,1</sup>, Xiaocong LIU <sup>a</sup>, Ting HE <sup>a</sup> and Jin GOU <sup>a</sup>

<sup>a</sup>College of Computer Science and Technology, Huaqiao University, Xiamen 36102, China

**Abstract.** In view of the non-interpretability of disease diagnosis models based on deep learning, a knowledge reasoning model based on medical knowledge graph for intelligent diagnosis is proposed. Given the patient symptom set, the co-occurrence of the patient and the disease is calculated, then the patient suffering from one disease is calculated. Based on the dynamic threshold value, the final disease diagnosis result of the patient is outputted. According to the symptoms of patients and the symptoms in the knowledge graph, the causal reasoning of the disease diagnosis is interpretable. Experiments on 145,712 pediatric electronic medical records in Chinese show that the proposed model can predict diseases with interpretability, and the accuracy reaches 82.12%.

**Keywords.** Knowledge graph, knowledge reasoning, disease diagnosis, Causal reasoning

## 1. Introduction

With the development of artificial intelligence technology, disease intelligent diagnosis based on mass medical data has become an important research content in the field of medical information [1]. Medical data, including electronic medical records (EMR), medical dictionaries, medical guidelines and genetic data, mainly exist in the form of text. In clinical practice, the diagnosis of diseases is usually based on patients' symptoms and examination results, combined with doctors' expertise, which is the most important decision support. Therefore, when using artificial intelligence technology to diagnose disease, it is natural to think of building medical professional knowledge into a medical knowledge graph to support medical decision-making [2,3]. The knowledge graph represents the entities and their relationships in a network structure, which can more clearly and intuitively mine the hidden knowledge [4]. In recent years, the research of disease diagnosis based on medical knowledge graph reasoning has been a hot spot in the application of medical artificial intelligence.

According to research, the methods of disease reasoning can be divided into two categories. The first method uses neural network to learn the representations of knowledge graph, which are used for realizing knowledge reasoning [5]. This method can automatically learn the feature of the data without the participation of experts, but because of the lack of interpretability, it is difficult to be widely used in clinical practice.

---

<sup>1</sup> Corresponding author: Huazhen Wang, College of Computer Science and Technology, Huaqiao University, Xiamen 36102, China; E-mail: wanghuazhen@hqu.edu.cn.

The second one is based on rules and logic, which gives comprehensive rules and ontology to obtain the result of knowledge reasoning, which is essentially logical knowledge reasoning.

Intelligent disease diagnosis is a knowledge-intensive application, and expert knowledge plays a key role in diagnosis decision. Therefore, making full use of the medical knowledge graph and logical reasoning method can obtain the disease reasoning results with high interpretability and accuracy. In this paper, we constructed a knowledge reasoning model based on causal reasoning for disease diagnosis task on real EMR datasets.

## 2. Related Work

### 2.1 Knowledge Reasoning Over Knowledge Graph

Knowledge reasoning is mainly focused on analyzing data and finding or inducing knowledge. In the aspect of knowledge reasoning, the most useful method is logical knowledge reasoning and the method based on neural network [6]. In terms of knowledge, there are many forms of knowledge, such as traditional tables or triples of knowledge graph. Knowledge graph, as a semantic network to reveal the relationship between entities, can formally describe the things and their relationships in the real world. discovering the connections between different things. Using knowledge graph as the source of knowledge reasoning can better assist intelligent system to discover knowledge.

With the emergence of knowledge graph, knowledge reasoning over knowledge graph has attracted more and more attention for serving intelligent system [7]. In the field of medicine, there are many medical knowledge graphs, such as Bio2RDF and Liked Life Data [8]. Shi et al. [9] constructed a graph knowledge from medical texts, and carried out semantic reasoning in this graph knowledge to realize disease diagnosis. But this method ignores the causal relationship between disease and symptoms, which plays an important role in disease diagnosis. Therefore, on the basis of knowledge reasoning, we apply causal reasoning method to the intelligent diagnosis of disease. Specifically, according to the symptoms of patients and the symptoms in the knowledge graph, the disease of patients can be inferred.

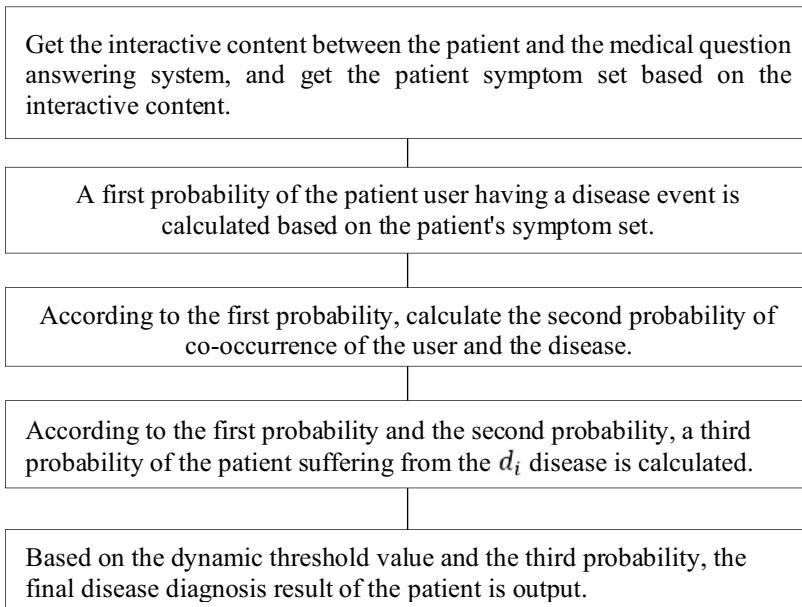
### 2.2 Intelligent Diagnosis of Related Diseases

The emphasis of computer-aided diagnosis is to use information technology to assist disease diagnosis [10]. Early researchers mainly relied on the diagnostic rules made by medical experts for auxiliary diagnosis. With the development of machine learning technology, researchers used Naive Bayesian algorithm and Support Vector Machine algorithm to build disease diagnosis model [11, 12]. In recent years, deep learning methods such as convolutional neural network (CNN) and RNN have been paid more and more attention by related researchers [13, 14]. Dong et al. [15] integrated the information of medical knowledge graph on the basis of deep learning method, so as to improve the accuracy of diseases diagnosis. However, due to lack of interpretability, these methods cannot be widely applied to clinical tasks. In this work, we propose a knowledge reasoning model based on medical knowledge and causal reasoning, trying to provide high interpretability and accuracy.

### 3. The Proposed Model for Disease Diagnosis

In this paper, we present a disease diagnosis model based on knowledge graph and causal reasoning, that is, according to the symptoms of patients and the symptoms in the knowledge graph, the disease of patients can be deduced, which is interpretable.

The essence of disease is symptoms, and symptoms are a phenomenon of disease. Patients may suffer from multiple diseases at the same time, and different diseases may not occur independently. The symptoms caused by the disease may not all show up. According to the patient's symptom set  $suArray$ , we can retrieve the symptoms  $svArray$  that do not appear on the patient but exist in the medical knowledge graph. According to the patient's symptoms  $suArray$ , we can retrieve the diseases associated with them in the knowledge graph, as well as the symptoms  $suArray$  of these diseases that do not appear in patients but exist in the medical knowledge graph. Then  $suArray$  and  $svArray$  were used to infer the patient's diseases. The implementation roadmap of the model is shown in Figure 1.



**Figure 1.** Implementation roadmap of the proposed model.

Let the set of diseases be  $D = \{d_1, d_2, \dots, d_i, \dots, d_n\}$ ,  $1 \ll i \ll n$ . Let the set of symptoms be  $S = \{s_1, s_2, \dots, s_j, \dots, s_m\}$ ,  $1 \ll j \ll m$ . The event  $\mathcal{D}_i$  is that the diseases in the collection  $D_i$  occurs simultaneously and none of the disease in set  $D - D_i$ , where  $D_i$  is a subset of the disease set  $D$ , and  $D_i$  is any one of  $2^D$ , that is  $D_i \in \{D_0, D_1, \dots, D_{2^n-1}\}$ . At first, the first probability  $P(S)$  that patient  $S$  has event  $\mathcal{D}_i$  is calculated, and its formula is:

$$P(S) = \sum_{k=0}^{2^n-1} P(\mathcal{D}_i)P(S|\mathcal{D}_i) \quad (1)$$

When  $P(\mathcal{S})$  is expanded according to the full probability formula, since  $d_1, d_2, d_3, \dots, d_n$  is not mutually exclusive events, it cannot constitute  $P(\mathcal{S})$  complete event graph. Set  $\mathcal{D}' \subset D$  makes the diseases in  $\mathcal{D}'$  constitute  $P(\mathcal{S})$  event graph.  $P(\mathcal{D}')$  is the probability of the complete event graph  $\mathcal{D}'$ , and  $P(\mathcal{S}|\mathcal{D}')$  is the probability of the complete event graph  $\mathcal{D}'$  under patient  $\mathcal{S}$ . Then, the second probability  $P(d_i|\mathcal{S})$  of patient  $\mathcal{S}$  and disease  $d_i$  is calculated, and the formula is:

$$\begin{aligned} P(d_i|\mathcal{S}) &= P\left(\left(\bigcup_{d_i \in \mathcal{D}'} \mathcal{D}'\right) \mathcal{S}\right) = P\left(\bigcup_{d_i \in \mathcal{D}'} \mathcal{D}' \mathcal{S}\right) = \sum_{d_i \in \mathcal{D}'} P(\mathcal{D}' \mathcal{S}) \\ &= \sum_{d_i \in \mathcal{D}'} P(\mathcal{D}') P(\mathcal{S}|\mathcal{D}') \end{aligned} \quad (2)$$

Obtaining the probability  $P(\mathcal{S})$  of the existence of the patient  $\mathcal{S}$  and the probability  $P(d_i|\mathcal{S})$  of the co-occurrence of the patient  $\mathcal{S}$  and the  $d_i$  disease, and calculating the third probability  $P(d_i|\mathcal{S})$  of the patient  $\mathcal{S}$  suffering from the  $d_i$  disease, wherein the formula is as follows:

$$P(d_i|\mathcal{S}) = \frac{P(d_i|\mathcal{S})}{P(\mathcal{S})} = \frac{\sum_{d_i \in \mathcal{D}'} P(\mathcal{D}') P(\mathcal{S}|\mathcal{D}')}{\sum_{k=1}^{2^n-1} P(\mathcal{D}_i) P(\mathcal{S}|\mathcal{D}_i)} \quad (3)$$

In the next part, a final disease diagnosis result of the patient  $\mathcal{S}$  is output based on the dynamic threshold and the third probability. The specific steps are as follows:

First of all, arrange the diseases in descending order according to the probability value obtained previously, to obtain the descending disease probability list  $O_r$  with the probability greater than 0, that is,  $Q_r = [disease\ 1: probability\ 1, disease\ 2: probability\ 2, \dots]$ ;

Then, the output threshold  $\lambda$  is set, and the probability of the descending disease probability list is compared with the set threshold  $\lambda$ . The disease whose probability is greater than or equal to the set threshold is constructed into the selected disease list  $O_{re}$  and output directly. The disease whose probability is less than the set threshold is constructed into the unselected disease list  $O_{rue}$  for the next step of screening.

According to  $\mathcal{D}'$ ,  $P(\mathcal{D}')$  and the probability  $P(s_j|\mathcal{D}')$  of symptom  $s_j$  occurring under the condition of complete event graph  $\mathcal{D}'$ , Calculate the fourth probability  $P(s_j)$  of the occurrence of the unmanifested symptom  $s_j$ , and its formula is:

$$P(s_j) = \sum_{\mathcal{D}' \subset D} P(\mathcal{D}') P(s_j|\mathcal{D}') \quad (4)$$

Selecting the non-appearing symptom corresponding to the maximum probability as the most likely appearing symptom to inquire about the patient; If the emerging symptom occurs, the emerging symptom is added to the emerging symptom set of the patient  $\mathcal{S}$ . The descending disease probability list is recalculated. And then, according to the probability of the occurrence of the disease  $d_i$  and the increment  $\Delta P(d_i)$  of the

probability of the occurrence of the disease  $d_i$  before and after interaction, calculating the increment  $\Delta I$  of the uncertainty, wherein the formula is as follows:

$$\Delta I = \sum_{d_i \in D \wedge P(d_i) > 0} \Delta I(d_i) = - \sum_{d_i \in D \wedge P(d_i) > 0} \log_2 \frac{P(d_i) + \Delta P(d_i)}{P(d_i)} \quad (5)$$

Then, calculate a dynamically adjusted threshold value  $\lambda' = \lambda - \Delta \lambda = \lambda - \Delta I$ , selected diseases in that unselected disease list  $Q_{rue}$  with the probability larger than the current threshold value, and construct an interactive disease list  $Q_{rei}$ ;

Finally, the selected disease list  $Q_{re}$  and the interactive disease list  $Q_{rei}$  are summarized to form a final disease diagnosis result list  $Q_{ree}$  as the final disease diagnosis result.

## 4. Experimental Setup and Results

### 4.1 Construction of Medical Knowledge Graph

In this paper, a bottom-up method was adopted to construct the medical knowledge graph. Firstly, we extracted the triplet of "disease" and "symptom" from the several medical websites: <http://tag.120ask.com/>, <http://jbk.39.net/>, <http://jbk.familydoctor.com.cn/>. Secondly, after entity alignment and knowledge fusion, these triples were stored in neo4j database. The final knowledge graph includes 7958 diseases, 5747 symptoms, and 65530 connections between them. An example diagram of the medical knowledge graph is shown in Figure 2.

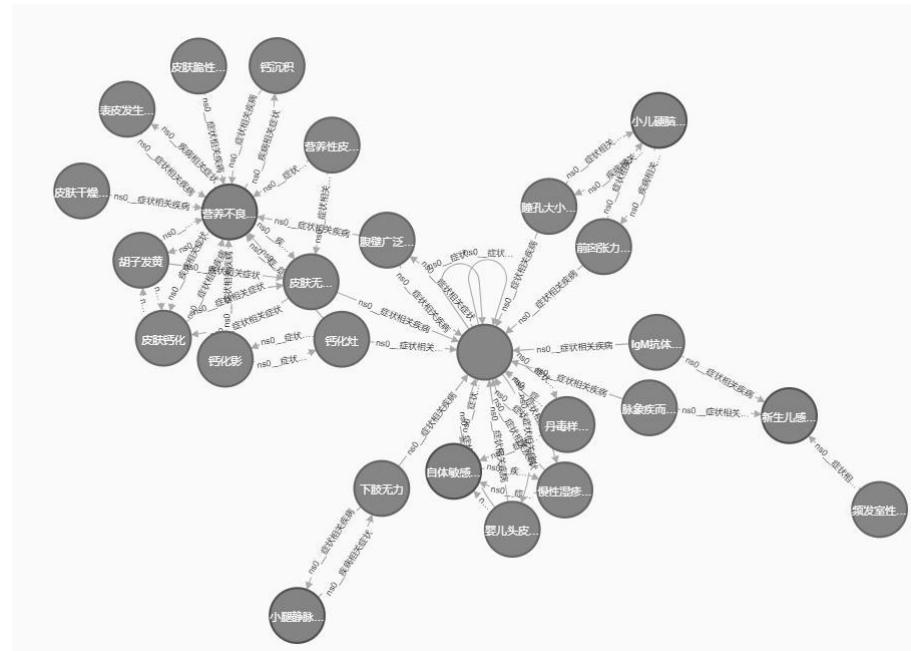


Figure 2. An example diagram of our medical knowledge graph.

## 4.2 Experimental Results

The dataset used in this paper come from a tertiary hospital, including 145,712 pediatric outpatient medical records, each of which has only one preliminary diagnosis. The purpose of this paper is to input the patient's symptoms and predict the patient's disease. The disease with the highest prediction probability is taken as the result. The evaluation index of the model is calculated as follows:

$$\text{Accuracy} = \frac{I_r}{I} \quad (6)$$

where  $I_r$  is the number of samples with correct diagnosis predicted by model, and  $I$  is the number of all samples.

We set the model proposed by Li et al. [16] as baseline model, which used deep convolutional neural network for pediatric disease prediction. Table 1 shows the experimental results.

**Table 1.** Comparison of experimental results between two models

Model	Accuracy
Our model	82.12%
[16]	71.07%

It can be seen from Table 1 that the accuracy of our model reaches 82.12%, which is 11.05% higher than that of CNN model, indicating that the causal reasoning model based on knowledge graph has outstanding performance in disease diagnosis.

## 5. Conclusion

Disease diagnosis is a prominent clinical application in clinical decision support. In this paper, we propose a knowledge reasoning model based on medical knowledge graph and causal reasoning for intelligent diagnosis. According to the patient's symptoms, we can retrieve the diseases associated with them in the knowledge graph, as well as the symptoms of these diseases that do not appear in patients but exist in the medical knowledge graph, thus inferring the patient's diseases. Then, our model can deduce the patient's disease set based on causal reasoning. Experiments on 145,712 pediatric electronic medical records in Chinese show that the proposed model can predict diseases with interpretability and high accuracy. It can be also applied to medical question and answer systems, medical auxiliary diagnosis and intelligent personal health assistants. Our future work will establish a more complete medical knowledge graph and apply our model to more diseases.

## Acknowledgments

Research works in this paper are supported by the National Key Technology R&D Program of China (No.2018YFB1402500), the Social Science Planning Foundation of Fujian Province (FJ2020B0033), and Huaqiao University's Academic Project Supported by the Fundamental Research Funds for the Central Universities (TZYB-202005).

## References

- [1] Sun XX. The Construction of academic knowledge mapping based on latent semantic analysis. Central China Normal University, 2013.
- [2] Zhao J, Liu K, He SZ, et al. knowledge Graph. Chin. J. Info., 2020;34 (09):111.
- [3] Giger ML, Chan HP, Boone J. Anniversary paper: history and status of CAD and quantitative image analysis: the role of medical physics and AAPM. *Medi Phys*, 2008; 35(12):5799-5820.
- [4] Dai Y, Wang S, Xiong NN, et al. A Survey on Knowledge Graph Embedding: Approaches, Applications and Benchmarks. *Electronics*, 2020, 9(5):750.
- [5] Wang D, Amriljaharadak A, Xiao Y. Dynamic knowledge inference based on bayesian network learning. *Mathematical Problems in Engineering*, 2020, 2020(4):1-9.
- [6] Guan SP, Jin XL, Jia YT, et al. Research progress of knowledge reasoning for knowledge graph. *Acta Software Sinica*. 2018, 29(10):74-102.
- [7] Hu GY. Research on the construction of knowledge graph of professional knowledge and skill system. *Industry and Info. Edu.* 2020(12): 123-127.
- [8] Jia LR, Liu J, Yu T, et al. Construction of traditional Chinese medicine knowledge graph. *J. Med. Info.* 2015, 36(08):51-53+59.
- [9] Shi L, Li S, Yang X, et al. Semantic health knowledge graph: semantic integration of heterogeneous medical knowledge and services. *Biomed Res Int*. 2017(2017):2858423.
- [10] Shortliffe EH, Axline SG, Buchanan BG, et al. An Artificial Intelligence program to advise physicians regarding antimicrobial therapy. *Compute Biomed Res*, 1973, 6(6):544-560.
- [11] Baati K, Hamdani T M, Alimi AM. Diagnosis of lymphatic diseases using a naive bayes style possibilistic classifier. *Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2013.
- [12] Kunwar V, Chandel K, Sabitha AS, et al. Chronic kidney disease analysis using data mining classification techniques. *Cloud Syst. Big Data Engineering*. IEEE, 2016.
- [13] Tsehay YK, Lay NS, Roth HR, et al. Depth learning architecture based on convolutional neural network for detecting prostate cancer on multi-parametric magnetic resonance images. *J. Medical Imaging 2017: Computer-Assisted Diagnosis*, Orlando, February 11-16, 2017. San Francisco: SPIE, 2017: 1013405.
- [14] Esteva A, Kuprel B, Novoa R A, et al. Corrigendum: Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 2017, 546(7660):686-686.
- [15] Dong LL, Cheng J, Zhang X. Research on disease diagnosis method integrating knowledge graph and deep learning. *Computer Sci. and Explor.* 2020, 14(05):815-824.
- [16] Li XZ, Wang HZ, Xiong YJ, et al. Application of convolutional neural network in pediatric disease prediction. *China Digital Med.* 2018,13(10): 11-13.

# The Larger the Better: Analysis of a Scalable Spectral Clustering Algorithm with Cosine Similarity

Guangliang CHEN<sup>a,1</sup>

<sup>a</sup>*San José State University, San José, California, United States*

**Abstract.** Chen (2018) proposed a scalable spectral clustering algorithm for cosine similarity to handle the task of clustering large data sets. It runs extremely fast, with a linear complexity in the size of the data, and achieves state of the art accuracy. This paper conducts perturbation analysis of the algorithm to understand the effect of discarding a perturbation term in an eigendecomposition step. Our results show that the accuracy of the approximation by the scalable algorithm depends on the connectivity of the clusters, their separation and sizes, and is especially accurate for large data sets.

**Keywords.** spectral clustering, cosine similarity, perturbation analysis

## 1. Introduction

Spectral clustering [1,2] was introduced at the beginning of the century as a very effective clustering approach. Given a set of objects  $O = \{o_1, o_2, \dots, o_n\}$  (such as images or documents) and a notion of similarity  $s(\cdot, \cdot)$  (e.g., Gaussian or cosine similarity), the first step of spectral clustering, as in other graph-based applications [3,4], is to construct a weighted graph  $\mathcal{G}$  from the given data using the similarity function  $s$ ,

$$\mathcal{G} = \{V, E, \mathbf{W}\}, \quad \mathbf{W} = (w_{ij}), \quad w_{ij} = s(o_i, o_j),$$

where  $V = O$  is the vertex set,  $E$  the edge set (there is an edge between objects  $o_i, o_j$  if and only if  $w_{ij} > 0$ ), and  $\mathbf{W}$  the weight matrix. Next, spectral clustering computes the eigenvectors of a normalized version of  $\mathbf{W}$  to obtain a low dimensional embedding of the data. Lastly, simple clustering algorithms like  $k$ -means are applied to the low dimensional coordinates to effectively cluster the given data. See Algorithm 1 for the Ng-Jordan-Weiss (NJW) version of spectral clustering [2].

---

<sup>1</sup>Corresponding Author: Department of Mathematics and Statistics, MH 308, San José State University, One Washington Square, San José, CA 95192-0103, United States; E-mail: guangliang.chen@sjsu.edu.

**Algorithm 1** Spectral Clustering (NJW)**Input:** Graph  $\mathcal{G} = \{V, E, \mathbf{W}\}$ , number of clusters  $k$ **Output:** A partition of the vertices in  $V$  into  $k$  clusters

- 1: Construct a diagonal degree matrix  $\mathbf{D}$  with  $\mathbf{D}_{ii} = \sum_j w_{ij}$ , and use it to normalize  $\mathbf{W}$  to obtain  $\widetilde{\mathbf{W}} = \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2}$ .
- 2: Find the top  $k$  eigenvectors  $\mathbf{v}_1, \dots, \mathbf{v}_k$  of  $\widetilde{\mathbf{W}}$  (corresponding to the largest  $k$  eigenvalues) and stack them as columns into a matrix  $\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_k]$ .
- 3: Group the row vectors of  $\mathbf{V}$  into  $k$  clusters by using the  $k$ -means algorithm.

Spectral clustering is a nonlinear clustering method due to the eigenvectors embedding step. As a result, it can easily handle non-convex geometries and accurately separate non-intersecting shapes. Spectral clustering has been successfully used in many applications, such as document clustering and image segmentation. However, spectral clustering is very slow on large data sets because of its high computational complexity associated to the  $n \times n$  matrix  $\mathbf{W}$ , which requires  $O(n^2)$  memory and  $O(n^3)$  time (for performing eigenvalue decomposition). Consequently, there has been considerable effort in the machine learning and data mining communities to develop fast, approximate spectral clustering algorithms that are scalable to large data [5,6,7,8,9,10,11,12,13,14,15,16,17]. Among those published methods, many use a “sampling plus extension” strategy by first working on a small number of landmark points selected from the given data and later extending the result to the full data set. As a result, the quality of the selected landmark points is crucial for the clustering accuracy and effective sampling can be very challenging in the setting of large, complex data sets.

## 2. A Review of the Scalable Spectral Clustering Algorithm

In [13] we tried to use the entire data set for direct clustering in the special setting of spectral clustering with cosine similarity:

$$\mathbf{W} = \mathbf{X}\mathbf{X}^T - \mathbf{I}, \quad (1)$$

where  $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_n]^T \in \mathbb{R}^{n \times d}$  represents a data matrix consisting of  $n$  unit vectors<sup>2</sup> in  $\mathbb{R}^d$  and  $\mathbf{I}$  is the identity matrix. To speed up spectral clustering on large data, we exploited the product form of the weight matrix  $\mathbf{W}$  for efficient implementation of spectral clustering.

Specifically, we assumed that the given data, despite its large size ( $n$ ), has some sort of low-dimensional structure in one of the following ways:

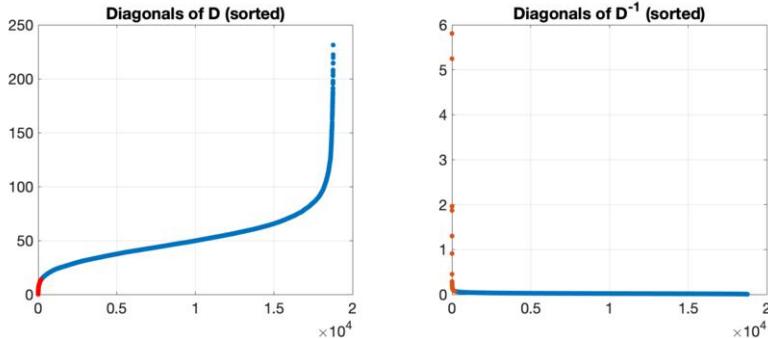
- (a) **The dimension  $d$  is also large but  $\mathbf{X}$  is sparse.** This assumption is often true for document data sets that are represented as document-term frequency matrices under the bag-of-words model [18]. An example is the well-known 20 newsgroups data set<sup>3</sup> stored in matrix format:  $\mathbf{X} \in \mathbb{R}^{n \times d}$  where  $n = 18,774$  (documents), and  $d = 61,188$  (terms), but the average row sparsity is only 129.7 (that is, on average, each document only contains 129.7 distinct words).

<sup>2</sup>The original data are vectors in  $\mathbb{R}^d$ . They have been normalized to have unit length in order to compute the cosine similarity.

<sup>3</sup>Available at <http://qwone.com/~jason/20Newsgroups/>.

- (b)  $d \ll n$  ( $\mathbf{X}$  does not need to be sparse). This assumption is true for many image data sets, such as the MNIST handwritten digits<sup>4</sup> ( $n = 70,000, d = 784$ ).

Note that for high dimensional non-sparse data, one can always use principal component analysis (PCA) to embed the given data into a few hundred dimensions (such that  $d \ll n$ ), which often prove to be sufficient.



**Figure 1.** Sorted diagonals of  $\mathbf{D}$  and  $\mathbf{D}^{-1}$  corresponding to the 20newsgroups data. The red part of each curve corresponds to the 1% of the data with the lowest degrees. The right plot shows that after that part is removed from the data, the remaining diagonals of  $\mathbf{D}^{-1}$  are approximately constant (mean: 0.0229, standard deviation: 0.0102).

We then showed that in those scenarios, spectral clustering with cosine similarity may be performed directly through efficient operations on the data matrix  $\mathbf{X}$  such as element-wise manipulation, matrix-vector multiplication, and low-rank singular value decomposition (SVD)<sup>5</sup>, thus completely avoiding the  $n \times n$  weight matrix  $\mathbf{W}$ :

- (1) The diagonal degree matrix  $\mathbf{D}$  is directly computed through matrix-vector multiplication

$$\mathbf{D} = \text{diag}((\mathbf{X}\mathbf{X}^T - \mathbf{I})\mathbf{1}) = \text{diag}(\mathbf{X}(\mathbf{X}^T\mathbf{1}) - \mathbf{1}), \quad (2)$$

where  $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^n$  is a constant vector. The diagonals of  $\mathbf{D}$  (degrees) are measures of the connectivity of the vertices in the graph. Afterwards, for some small  $\alpha > 0$ , a fraction  $\alpha$  of the data in  $\mathbf{X}$  with the lowest degrees is discarded as outliers such that for the remaining data (still denoted by  $\mathbf{X}$ ), the diagonal of  $\mathbf{D}^{-1}$  is approximately constant. See Figure 1 for a demonstration.

- (2) For the correspondingly reduced weight matrix (still denoted by  $\mathbf{W}$ ), write the normalized cosine similarity matrix  $\tilde{\mathbf{W}}$  as follows:

$$\tilde{\mathbf{W}} = \mathbf{D}^{-1/2} \underbrace{(\mathbf{X}\mathbf{X}^T - \mathbf{I})}_{\mathbf{W}} \mathbf{D}^{-1/2} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^T - \mathbf{D}^{-1}, \quad \tilde{\mathbf{X}} = \mathbf{D}^{-1/2}\mathbf{X}. \quad (3)$$

Disregard the  $\mathbf{D}^{-1}$  term in (3) to use the left singular vectors of  $\tilde{\mathbf{X}}$  to approximate the corresponding eigenvectors of  $\tilde{\mathbf{W}}$  (note that  $\tilde{\mathbf{X}}$  is sparse or low dimensional, dependent on  $\mathbf{X}$ , and thus its SVD can be computed efficiently).

<sup>4</sup>Available at <http://yann.lecun.com/exdb/mnist/>.

<sup>5</sup>All of these operations have a linear complexity in  $n$ , the number of data points.

See Algorithm 2 for the full algorithm.

**Algorithm 2** Scalable spectral clustering with cosine similarity

**Input:** Data matrix  $\mathbf{X}$  (sparse or of moderate dimension, with  $L_2$  normalized rows), #clusters  $k$ , fraction of data to be removed  $\alpha$

**Output:** Clusters  $C_1, \dots, C_k$  and a set of outliers  $C_0$

**Steps:**

- 1: Compute the degree matrix  $\mathbf{D}$  via (2) and label the bottom  $(100\alpha)\%$  points as outliers (stored in the set  $C_0$ ). Remove  $C_0$  from the input data.
- 2: Calculate  $\tilde{\mathbf{X}} = \mathbf{D}^{-1/2}\mathbf{X}$  for the remaining data and find its top  $k$  left singular vectors  $\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_k \in \mathbb{R}^n$  by rank- $k$  SVD. Let  $\tilde{\mathbf{U}}_k = [\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_k] \in \mathbb{R}^{n \times k}$ .
- 3: Normalize the rows of  $\tilde{\mathbf{U}}_k$  to have unit length and apply  $k$ -means to find  $k$  clusters  $C_1, \dots, C_k$ .

We tested our scalable algorithm for clustering large text and image data sets and obtained comparable accuracy with the plain implementation but our algorithm runs much faster [13]. Recently, we have successfully extended the work to deal with general similarity functions (such as Gaussian) [15,17].

### 3. Analysis

In this section we conduct careful and rigorous analysis of the effect of the term  $\mathbf{D}^{-1}$  in (3) on the eigenvectors of  $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$ .

#### 3.1. Insights

We start by making the following observations:

- The matrix  $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$  in (3) is symmetric and positive semidefinite, and  $\mathbf{D}^{-1}$  can be viewed as a perturbation to it. Thus, the research conducted here is along the direction of perturbation analysis of the eigenspace of a positive semidefinite matrix.
- The matrix  $\tilde{\mathbf{W}}$  is similar to a row-stochastic matrix  $\mathbf{P} = \mathbf{D}^{-1}\mathbf{W}$ :

$$\tilde{\mathbf{W}} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} = \mathbf{D}^{1/2}\mathbf{P}\mathbf{D}^{-1/2} \quad (4)$$

Therefore, the largest eigenvalue of  $\tilde{\mathbf{W}} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^T - \mathbf{D}^{-1}$  is 1. The next  $k-1$  largest eigenvalues of  $\tilde{\mathbf{W}}$  are expected to be close to 1 and meanwhile, there should be a significant drop at the  $(k+1)$ -th eigenvalue.

- If  $\mathbf{D}$  has a constant diagonal, then  $\tilde{\mathbf{W}}$  have the same eigenvectors with  $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$  (but not the same eigenvalues), thus discarding  $\mathbf{D}^{-1}$  won't change the eigenvectors in such a case.
- Adding a constant multiple of the identity matrix to  $\tilde{\mathbf{W}}$  does not change its eigenvectors (but only shifts its eigenvalues by  $\beta$ ):

$$\tilde{\mathbf{W}} + \beta\mathbf{I} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^T + (\beta\mathbf{I} - \mathbf{D}^{-1}) \quad (5)$$

We will select  $\beta$  such that  $\beta\mathbf{I} - \mathbf{D}^{-1}$  is as small as possible (with respect to appropriate matrix norm) while being positive semidefinite.

- The underlying function of  $\mathbf{D}^{-1}$  is  $f(x) = 1/x$  which flattens out quickly as  $x$  increases. This implies that  $\mathbf{D}^{-1}$  is often close to being constant diagonal for large data sets (in which case, most diagonals of  $\mathbf{D}$  are expected to be large). We will conduct analysis to estimate the magnitude of the diagonals of  $\mathbf{D}$  in such settings.

### 3.2. Analysis

Let  $\tilde{\mathbf{U}}_k \in \mathbb{R}^{n \times k}$  be the matrix consisting of the top  $k$  eigenvectors of  $\tilde{\mathbf{W}}$ , which is used by the exact Ng-Jordan-Weiss spectral clustering algorithm. Denote by  $\hat{\mathbf{U}}_k \in \mathbb{R}^{n \times k}$  the matrix consisting of the top  $k$  eigenvectors of  $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$  (which are also left singular vectors of  $\tilde{\mathbf{X}}$ ), which is used by the scalable spectral clustering algorithm as an approximation to  $\tilde{\mathbf{U}}_k$ . Our goal here is to relate the Grassmann distance between  $\hat{\mathbf{U}}_k$  and  $\tilde{\mathbf{U}}_k$ , i.e.,  $\|\hat{\mathbf{U}}_k \hat{\mathbf{U}}_k^T - \tilde{\mathbf{U}}_k \tilde{\mathbf{U}}_k^T\|_F$ , to the perturbation term  $\mathbf{D}^{-1}$ .

Using [19, Theorem A.1] we can prove the following result.

*Theorem 3.1.* Let  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$  be the eigenvalues of  $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$ , and define  $\delta_k = \lambda_k - \lambda_{k+1} > 0$ . Then

$$\|\tilde{\mathbf{U}}_k \tilde{\mathbf{U}}_k^T - \hat{\mathbf{U}}_k \hat{\mathbf{U}}_k^T\|_F \leq \frac{2\sqrt{n}}{\delta_k} \|\mathbf{D}^{-1}\|_2. \quad (6)$$

*Proof.* Write the degree matrix as

$$\mathbf{D} = \text{diag}(d_1, \dots, d_n), \quad 0 < d_1 \leq \dots \leq d_n$$

Let  $\beta = d_1^{-1}$ , and define

$$\tilde{\mathbf{W}}^{(\beta)} = \tilde{\mathbf{W}} + \beta\mathbf{I} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^T + (\beta\mathbf{I} - \mathbf{D}^{-1})$$

Clearly, both  $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$  and  $\beta\mathbf{I} - \mathbf{D}^{-1}$  are symmetric and positive semidefinite. It follows that  $\tilde{\mathbf{W}}^{(\beta)}$  is also positive semidefinite. Additionally, the eigenvectors of  $\tilde{\mathbf{W}}^{(\beta)}$  are the same with those of  $\tilde{\mathbf{W}}$  (so the top  $k$  eigenvectors of  $\tilde{\mathbf{W}}^{(\beta)}$  are still  $\tilde{\mathbf{U}}_k$ ). Thus,  $\beta\mathbf{I} - \mathbf{D}^{-1}$  can be viewed as a positive semidefinite perturbation matrix to the positive semidefinite matrix  $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$ .

Using the perturbation result in [19, Theorem A.1], we obtain that

$$\|\tilde{\mathbf{U}}_k \tilde{\mathbf{U}}_k^T - \hat{\mathbf{U}}_k \hat{\mathbf{U}}_k^T\|_F \leq \frac{2}{\delta_k} \|\beta\mathbf{I} - \mathbf{D}^{-1}\|_F$$

provided that  $\|\beta\mathbf{I} - \mathbf{D}^{-1}\|_F \leq \delta_k/4$ .

Since

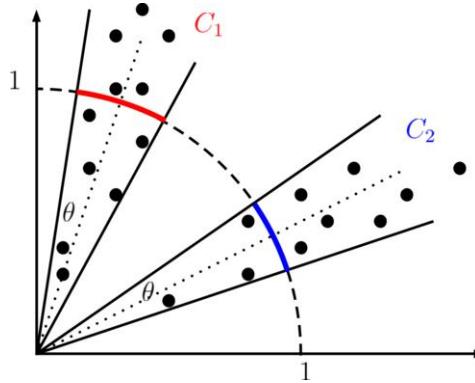
$$\|\beta\mathbf{I} - \mathbf{D}^{-1}\|_F^2 = \sum_{i=1}^n \left( \frac{1}{d_1} - \frac{1}{d_i} \right)^2 = \frac{1}{d_1^2} \sum_{i=1}^n \left( \frac{d_i - d_1}{d_i} \right)^2 \leq \frac{1}{d_1^2} \sum_{i=1}^n 1 = \frac{n}{d_1^2}$$

we have

$$\|\tilde{\mathbf{U}}_k \tilde{\mathbf{U}}_k^T - \hat{\mathbf{U}}_k \hat{\mathbf{U}}_k^T\|_F \leq \frac{2}{\delta_k} \cdot \sqrt{\frac{n}{d_1^2}} = \frac{2\sqrt{n}}{\delta_k} \|\mathbf{D}^{-1}\|_2.$$

This completes the proof.  $\square$

Next, we estimate the magnitude of the diagonals of  $\mathbf{D}$  in the setting of image or document data (where the data matrix has nonnegative entries). For this purpose, we need to assume that the data is a sample from a mixture distribution in the first orthant of  $\mathbb{R}^d$ . Specifically, we suppose that each cluster  $C_j$  is a random sample of  $n_j$  points from a cone that is within an angle of  $\theta$  around a unit vector  $\mathbf{t}_j$ , where  $n_j \geq \gamma n$  for some fixed constant  $\gamma > 0$ . See Figure 2 for an illustration. The data used in Algorithm 2,  $\mathbf{X} \in \mathbb{R}^{n \times d}$ , represent the normalized version of the sample.



**Figure 2.** Mixture of cones model (when  $k = 2$ ). Each cluster is a collection of samples from a cone concentrated around its axis within angle  $\theta$ . The original observations are projected onto the unit sphere (represented by the dashed curve) to have unit length for computing the cosine similarity used by Algorithm 2.

We can prove the following result on the magnitude of the degrees of the points.

*Theorem 3.2.* Under the above assumptions,

$$d_i \geq \gamma \cos^2 \theta - 1, \quad \text{for all } i = 1, \dots, n. \quad (7)$$

*Proof.* Consider the data point  $\mathbf{x}_i$  and suppose that it comes from  $C_j$ .

By (2), we have

$$d_i = \mathbf{x}_i^T (\mathbf{X}^T \mathbf{1}) - 1 = \mathbf{x}_i^T \left( \sum_{\ell=1}^n \mathbf{x}_\ell \right) - 1$$

Since all the data points are in the first quadrant,

$$d_i \geq \mathbf{x}_i^T \left( \sum_{\mathbf{x}_\ell \in C_j} \mathbf{x}_\ell \right) - 1 = \mathbf{x}_i^T (n_j \mathbf{m}_j) - 1$$

where

$$\mathbf{m}_j = \frac{1}{n_j} \sum_{\mathbf{x}_\ell \in C_j} \mathbf{x}_\ell$$

is the centroid of  $C_j$ .

Using elementary geometry, we have

$$\|\mathbf{m}_j\| \geq \cos \theta.$$

It follows that

$$\begin{aligned} d_i &\geq n_j (\mathbf{x}_i^T \mathbf{m}_j) - 1 \\ &\geq n_j \cdot (1 \cdot \|\mathbf{m}_j\| \cdot \cos \theta) - 1 \\ &\geq (\gamma n) \cdot \cos \theta \cdot \cos \theta - 1 \\ &= \gamma n \cos^2 \theta - 1 \end{aligned}$$

This completes the proof.  $\square$

Combining the above two theorems immediately give the following result.

*Corollary 3.1.* Under the same assumptions as in Theorems 3.1 and 3.2,

$$\|\tilde{\mathbf{U}}_k \tilde{\mathbf{U}}_k^T - \hat{\mathbf{U}}_k \hat{\mathbf{U}}_k^T\|_F \leq \frac{2\sqrt{n}}{\delta_k(\gamma n \cos^2 \theta - 1)}. \quad (8)$$

#### 4. Discussions

Theorem 3.1 shows that the closeness of  $\hat{\mathbf{U}}_k$  to  $\tilde{\mathbf{U}}_k$  is bounded by the  $k$ th eigengap  $\delta_k$  (inversely) and the spectral norm of the matrix  $\mathbf{D}$ . The size of eigengap  $\delta_k$  corresponds to the separation between the different clusters. If the  $k$  clusters are all well separated, then  $\delta_k$  will be considerably bigger than zero.

Theorem 3.2 shows that the magnitude of the degrees of the data is  $O(n)$ . For large data sets, the degrees of the points are also large, on the same order with the size of the data. As a result, the diagonals of  $\mathbf{D}^{-1}$  will be very small, and thus the perturbation to  $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$  will be small.

Theorem 3.1 combines Theorems 3.1 and 3.2 to relate the Grassmann distance of  $\hat{\mathbf{U}}_k$  to  $\tilde{\mathbf{U}}_k$  to the following quantities:

- $n$ : size of the data set.
- $\gamma$ : fraction of the smallest cluster in the data.
- $\theta$ : tightness of the clusters. The smaller  $\theta$  (and correspondingly the larger  $\cos^2 \theta$ ), the more connected within each cluster.
- $\delta_k$ : eigengap which measures separation among the clusters.

Overall, the Grassmann distance of  $\hat{\mathbf{U}}_k$  to  $\tilde{\mathbf{U}}_k$  is  $O(n^{-1/2})$ .

## 5. Conclusion

We showed through a matrix perturbation analysis that, for large data sets that have well connected clusters and sufficient separation between them, the scalable spectral clustering algorithm (Algorithm 2) provides a close approximation to the plain algorithm (Algorithm 1). The larger the data set, the better the approximation!

## References

- [1] Shi J, Malik J. Normalized cuts and image segmentation. *IEEE Trans Pattern Anal Mach Intell.* 2000;22(8):888–905.
- [2] Ng A, Jordan M, Weiss Y. On spectral clustering: analysis and an algorithm. In: *Advances in Neural Information Processing Systems 14*; 2001. p. 849–856.
- [3] Pietrabissa A, Celsi LR, Cimorelli F, Suraci V, Priscoli FD, Giorgio AD, et al. Lyapunov-based design of a distributed wardrop load-balancing algorithm with application to software-defined networking. *IEEE Transactions on Control Systems Technology.* 2019;27(5):1924–1936.
- [4] Celsi LR, Giorgio AD, Gambuti R, Tortorelli A, Priscoli FD. On the many-to-many carpooling problem in the context of multi-modal trip planning. In: *Proceedings of the 25th Mediterranean Conference on Control and Automation (MED)*; 2017. p. 303–309.
- [5] Fowlkes C, Belongie S, Chung F, Malik J. Spectral grouping using the Nyström method. *IEEE Trans Pattern Analysis and Machine Intelligence.* 2004;26(2):214–225.
- [6] Yan D, Huang L, Jordan M. Fast approximate spectral clustering. In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; 2009. p. 907–916.
- [7] Wang L, Leckie C, Ramamohanarao K, Bezdek J. Approximate spectral clustering. vol. 5476 of *Advances in Knowledge Discovery and Data Mining. PAKDD 2009, Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer; 2009.
- [8] Wang L, Leckie C, Kotagiri R, Bezdek J. Approximate pairwise clustering for large data sets via sampling plus extension. *Pattern Recognition.* 2011;44:222–235.
- [9] Tasdemir K. Vector quantization based approximate spectral clustering of large datasets. *Pattern Recognition.* 2012;45(8):3034–3044.
- [10] Choromanska A, Jebara T, Kim H, Mohan M, Monteleoni C. Fast spectral clustering via the Nyström method. vol. 8139 of *Algorithmic Learning Theory. ALT 2013. Lecture Notes in Computer Science*. Jain S, Munos R, Stephan F, Zeugmann T, editors. Berlin, Heidelberg: Springer; 2013.
- [11] Cai D, Chen X. Large scale spectral clustering via landmark-based sparse representation. *IEEE Transactions on Cybernetics.* 2015;45(8):1669–1680.
- [12] Moazzen Y, Tasdemir K. Sampling based approximate spectral clustering ensemble for partitioning data sets. In: *Proceedings of the 23rd International Conference on Pattern Recognition*; 2016. .
- [13] Chen G. Scalable spectral clustering with cosine similarity. In: *Proceedings of the 24th International Conference on Pattern Recognition (ICPR), Beijing, China*; 2018. .
- [14] Pham K, Chen G. Large-scale spectral clustering using diffusion coordinates on landmark-based bipartite graphs. In: *Proceedings of the 12th Workshop on Graph-based Natural Language Processing (TextGraphs-12)*. New Orleans, Louisiana: Association for Computational Linguistics; 2018. p. 28–37.
- [15] Chen G. A scalable spectral clustering algorithm based on landmark-embedding and cosine similarity. In: Bai X., Hancock E., Ho T., Wilson R., Biggio B., Robles-Kelly A. (eds) *Structural, Syntactic, and Statistical Pattern Recognition. S+SSPR 2018. Lecture Notes in Computer Science*. vol. 11004. Cham: Springer; 2018. .
- [16] Chen G. Matlab implementation details of a scalable spectral clustering algorithm with cosine similarity. In: *Proceedings of the 2nd Workshop on Reproducible Research in Pattern Recognition*. Cham: Springer; 2018. .
- [17] Chen G. A general framework for scalable spectral clustering based on document models. *Pattern Recognition Letters.* 2019;125:488–493.
- [18] Aggarwal C, Zhai C. In: *A survey of text clustering algorithms*. Boston, MA: Springer US; 2012. p. 77–128.
- [19] Chen G, Lerman G. Foundations of a multi-way spectral clustering framework for hybrid linear modeling. *Found Comput Math.* 2009;DOI 10.1007/s10208-009-9043-7.

# About Strong Dependence of the Complexity of Analysis of the Random 3-CNF Formulas on the Ratio of Number of Clauses to the Number of Variables

Sergey I. UVAROV<sup>a, 1</sup>

<sup>a</sup> Institute of control Science RAS, Moscow, Russia

**Abstract.** The results of a computational experiment on the assessment of the complexity of proving the unsatisfiability of random 3-CNF logical formulas are presented. The dependence of the complexity of this proving on the  $R$ -ratio of the number of clauses to the number of variables is demonstrated. The computational experiment was carried out for the range of the  $N$ -number of variables from 256 to 512. An exponential dependence of the median complexity of proving the unsatisfiability of formulas on the number of variables was revealed for each of  $R$  value: 4.3, 4.6, 5.0, 5.5, 6.0. A formula is constructed that approximates the results of the experiment. According to this formula the exponential component of the median complexity of the analysis of random 3-CNF is estimated as 2 to the power  $N / (8.4R - 17.8)$ .

**Keywords.** satisfiability (SAT) problem, conjunctive normal form (CNF), clause, variable, literal, complexity.

## 1. Introduction

The ultimate goal of the research is the question of whether the complexity of the SAT-problem is polynomial or exponential. In this paper, attention is focused on the relationship between the complexity of proving the unsatisfiability of random 3-CNF formulas and the  $R$ -ratio of the  $M$ -number of clauses to the  $N$ -number of variables.

Focusing on the proof of unsatisfiability is methodically justified by the fact that it makes necessary to consider all branches of the formula analysis tree. The number of branches in this tree analysis serves as a measure of complexity. At the same time, the process of proving the unsatisfiability of logical formulas is important for artificial intelligence systems. This is the base of automated theorem proving (also known as ATP or automated deduction). The complexity of automated deduction in general CNF formulas restrict the number of variables involved in deduction. Such limitation on the number of variables arouses interest in estimating the complexity of analyzing CNF formulas. and to the development of the fastest possible computational algorithms.

As an instances for a computational experiment are selected widely used random 3-CNF formulas [1-7]. In the course of the study, random 3-CNF formulas (with relatively small value of  $R$ ) demonstrated consistently high complexity (of the analysis).

---

<sup>1</sup> Corresponding author, Institute of control Science RAS, Moscow, Russia; E-mail: uvarov53@gmail.com.

Collected representative statistics show an exponential grows of complexity with an increase in the number  $N$  of variables from 256 to 512, and fixed values of  $R$  (4.3, 4.6, 5.0, 5.5, 6.0).

The use of the median characteristics of the experimental data made it possible to obtain predictable dependencies illustrated graphically.

The research was conducted using SAT-solver based on well-known Davis–Putnam–Logemann–Loveland (DPLL) algorithm [1,2] with *backtracking*, using *local search* [1, 2] and *learning clauses* [1-3] procedures.

We find strong dependence of the complexity of the 3-CNF problem on  $R$ . In the computational experiment the denominator of the exponent depends linearly on  $R$ .

The median complexity of analysis of random 3-CNF formulas from  $N$  variables is estimated as 2 to the power of  $N/(\beta R - \delta)$ .

Perhaps in the future it will be possible to prove that all of SAT-solvers of this kind (based on, backtracking, local search and learning clauses) are characterized by an exponential complexity estimate for arbitrary  $N$ .

## 2. Random 3-CNF

We use a well-known and very convenient method [1,2] for generating random 3-CNF formulas. 3-CNF formula is a conjunction of a set of  $M$  clauses. Each clause is a disjunction of three terms (literals) of different boolean variables. A term is a logical variable  $v$  itself or its negation  $\neg v$ . The total number of such three-term clauses for  $N$  variables is  $8 \binom{N}{3} \sim O(N^3)$ , and each clause is numbered.

When constructing a formula from this general set of numbered clauses, the required number  $M$  of clauses is selected randomly. 3-CNF is satisfiable if there is any logical variables assignment that gives true value to the formula.

Such a simple method for generating of random formulas has a remarkable property that limits the area of simultaneous existence of satisfiable and unsatisfiable formulas. It is proved [3,4], that with an increase in the number of variables at  $3.52 > R$ , unsatisfiable formulas are generated negligibly rarely. At  $R > 4.51$  negligibly rarely are generated satisfiable formulas. Statistical studies have shown that at  $4.3 < R$  the share of unsatisfiable formulas prevails over the share of satisfiable ones.

It is generally accepted that for randomly generated 3-CNF formulas with certain  $R$  ratio of clauses to variables it is difficult to analyze their satisfiability or unsatisfiability.

## 3. Experimental Results

Table 1 shows the results of a computational experiment to assess the complexity of proving the unsatisfiability of formulas.

In the cells of the table there are  $\mu E(N, R)$  median values of the complexity of proving the unsatisfiability of formulas expressed via the number of branches constructed. Further, in parentheses, the deviation of the  $\mu E(N, R)$  median value from the value specified by the obtained approximation formula is given, expressed as a percentage.

**Table 1.** The median value of the number of branches in the proof of the unsatisfiability of a random 3-CNF formula depending on  $N$  and  $R$ .

N	R*=4.3	R=4.6	R=5.0	R=5.5	R=6.0
256	408·10^2 (2%)	8960 (4%)	3010 (1%)		
288	152·10^3 (12%)	260·10^2 (4%)	7670 (1%)		
320	493·10^3 (7%)	799·10^2 (2%)	191·10^2 (1%)	5020 (3%)	
352	154·10^4 (26%)	217·10^3 (5%)	450·10^2 (7%)	110·10^2 (2%)	
384	493·10^4 (7%)	777·10^3 (3%)	145·10^3 (19%)	274·10^2 (13%)	7830 (2%)
416	195·10^5 (8%)	207·10^4 (8%)	310·10^3 (3%)	503·10^2 (2%)	151·10^2 (2%)
448	565·10^5 (7%)	564·10^4 (2%)	719·10^3 (5%)	112·10^3 (1%)	291·10^2 (4%)
480	193·10^6 (7%)	151·10^5 (8%)	179·10^4 (13%)	245·10^3 (1%)	593·10^2 (2%)
512		422·10^5 (12%)	458·10^4 (3%)	546·10^3 (3%)	118·10^3 (2%)

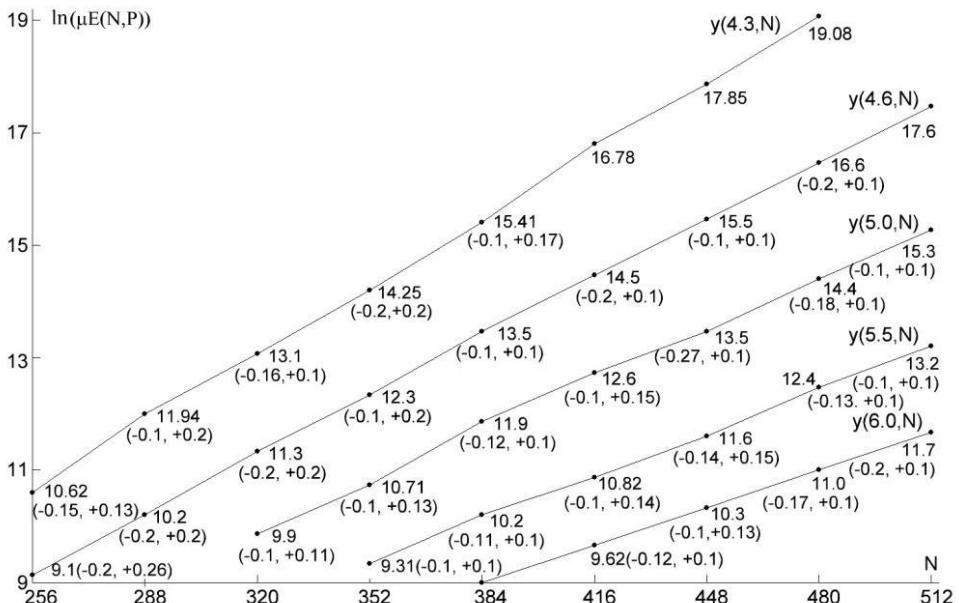
Columns 3 through 6 show the results of an experiment with randomly generated formulas where the number of clauses is a priori set by the value  $M = NR$ .

It should be noted that for  $N \geq 256$  and  $R \geq 4.6$ , no satisfiable formula has been obtained.

The second column represents random unsatisfiable formulas such that in their construction the number of clauses involved gradually increases until an unsatisfiable formula is obtained. Here, half of the random formulas are unsatisfiable with  $R < 4.3$ .

In this column, the value  $R^* = 4.3$  is the median value of the ratio of the number of clauses to the number of variables at which the proportion of unsatisfiable formulas becomes prevalent.  $R^*$  corresponds to the «phase transition» from satisfiability to unsatisfiability [1, 2] for random 3-CNF formulas.

The experimental results presented in Table 1 are graphically illustrated in Figure 1. Each point of the graph corresponds to the natural logarithm of the number  $\mu E(N, R)$  presented in the cell of Table 1.



**Figure 1.** The logarithmic dependence  $\mu E(N, R)$ .

In proving the unsatisfiability of each of the formulas, the value  $Ej(N, R)$ ,  $j = 1 \div K$  was calculated, which represents the number of branches in the analysis tree.

The number  $\mu E(N, R)$  is the median value among the complexities  $Ej(N, R)$ ,  $j = 1 \div K$ , which are associated with this cell of Table 1.

If  $K=64$  or more random formulas were constructed and analyzed for a cell of Table 1, the confidence interval for the probability 0.95 is given. Confidence interval is shown (in brackets) next to the value of the natural logarithm of  $\mu E(N, R)$  in Figure 1, The absence of a confidence interval means that the median complexity is obtained for  $K=16$  random formulas.

#### 4. Some Generalizations

The approximation by the least squares method of the experimental data presented in Figure 1 is expressed in the form:

$$\begin{aligned}y(4.3, N) &= N/26.18 + b(4.3), \\y(4.6, N) &= N/30.09 + b(4.6), \\y(5.0, N) &= N/34.93 + b(5.0), \\y(5.5, N) &= N/41.44 + b(5.5), \\y(6.0, N) &= N/46.65 + b(6.0).\end{aligned}$$

The coefficients  $b(4.3)$ ,  $b(4.6)$ ,  $b(5.0)$ ,  $b(5.5)$ , and  $b(6.0)$  are not of great interest in this context. It is convenient to introduce  $B(R) = e^{b(R)}$ . In a more visual form, the approximation of the exponential component of the estimation of complexity of the 3-CNF problem is expressed as:

$$\begin{aligned}\mu E(N, 4.3) &\approx B(4.3) \cdot 2^{N/18.15} = B(4.3) \cdot 2^{N/D(4.3)}, \\ \mu E(N, 4.6) &\approx B(4.6) \cdot 2^{N/20.86} = B(4.6) \cdot 2^{N/D(4.6)}, \\ \mu E(N, 5.0) &\approx B(5.0) \cdot 2^{N/24.20} = B(5.0) \cdot 2^{N/D(5.0)}, \\ \mu E(N, 5.5) &\approx B(5.5) \cdot 2^{N/28.72} = B(5.5) \cdot 2^{N/D(5.5)}, \\ \mu E(N, 6.0) &\approx B(6.0) \cdot 2^{N/32.35} = B(6.0) \cdot 2^{N/D(6.0)}.\end{aligned}$$

Where  $B(4.3)=2.267$ ,  $B(4.6)=1.896$ ,  $B(5.0)=2.020$ ,  $B(5.5)=2.283$  and  $B(6.0)=2.039$ .

Figure 2 illustrates the dependence of  $D$  the denominator in the exponent on  $R$ . Visually, all five points  $D(4.3)$ ,  $D(4.6)$ ,  $D(5.0)$ ,  $D(5.5)$  and  $D(6.0)$  lie almost in a straight line. This suggests a linear relationship between  $D$  and  $R$ .

Representing  $D(R)$  in the form  $D = \beta R - \delta$  and one more apply the least squares method, we obtain  $\beta = 8.4$ ,  $\delta = 17.8$ .

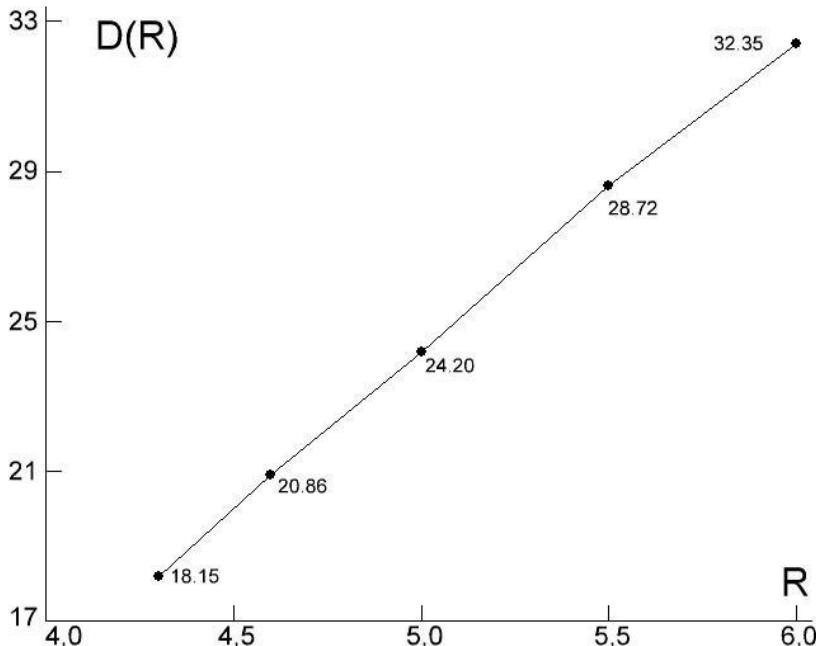
The median complexity (expressed in the number of branches of the analysis tree) of the proof of the unsatisfiability of 3-CNF formulas which are constructed using random numbers, obtained in the course of the computational experiment can be represented as:

$$\mu E(N, R) \approx B(R) \cdot 2^{N/(8.4R-17.8)}.$$

In an early publication [6] devoted to the study of random 3-CNF formulas, the median value of the denominator of exponent at  $R^{\sim}=4.3$  was estimated as  $D^{\sim}(4.3)=17$ , and for  $R^{\sim}=10$  the value  $D^{\sim}(10)=57$ .

This estimate can be interpreted so that the algorithm used in [6] for solving random 3-CNF problem is characterized by  $D^{\sim}=7R-13$ . Accordingly,  $\beta^{\sim}=7$ ,  $\delta^{\sim}=13$ .

The values of the  $\beta$  and  $\delta$  coefficients depend on the used procedures of *local search* and *learning clauses* and can be used to compare SAT-solvers.



**Figure 2.** Graphical representation of the dependence of the denominator  $D$  of the exponent on  $R$ .

## 5. Conclusions

The study confirmed the well-known hypothesis that the most difficult to analyze random 3-CNF formulas are concentrated in the area of coexistence of both satisfiable and unsatisfiable examples. This area is defined by the range  $3.52 \leq R \leq 4.51$ .

In the experiment carried out, the interval of  $R$  values where both satisfiable and unsatisfiable formulas are simultaneously present ranges from 4.01 to 4.56. The interval width is 0.55.

It is very likely that the study of formulas in this area is most fruitful for answering the question about the polynomial or exponential nature of the SAT problem.

The random formulas obtained by the described method [1, 2], [4-6] turned out to be rather complicated. Indeed, for formulas with  $N = 480$  associated with the second column of Table 1, the proof of the unsatisfiability requires the analysis of a tree containing more than 193,000,000 branches.

For a fairly advanced algorithm for proving the unsatisfiability of random 3-CNF formulas, in the range of the  $N$  number of variables from 256 to 512, the exponential nature of the dependence of the complexity of such a proof on  $N$  is revealed. This result correlates well with the numerous previous published results [1,2,6] obtained for CNF formulas (with fewer  $N$  number of variables).

Demonstrated strong dependence of the analysis complexity of formulas on  $R$ . This points the way for further improvement of the methods for producing formulas,

which are difficult to analyze. For example, in [7], a generator is proposed that balances the load on literals of variables by generating 3-CNF (partially random) formulas. For such generating the fraction of unsatisfiable formulas becomes predominant at  $R \approx 3.5$ .

As a result of a computational experiment, a linear dependence of the denominator of the exponent on  $R$ - the ratio of the number of clauses to the  $N$  number of variables.

In the research the exponential component of median complexity of unsatisfiability analysis of random 3-CNF formulas approximates by a function of  $N$  and  $R$  namely 2 to the power of  $N/(8.4R-17.8)$ , in the range  $N=256 \div 512$ ,  $R=4.3 \div 6.0$ .

In the future, it is planned to try to prove that when using SAT-solvers built on the basis of *backtracking* with *local search* and *learning clauses* procedures the complexity of SAT-problem is exponential.

## References

- [1] Biere A, Heule M, Maaren H, Walsh T. Handbook of Satisfiability. IOS Press; 2009. p. 1-966.
- [2] Gomes C, Rautz H, Sabharwal A, Selman B. Satisfiability Solvers. in Handbook of Knowledge Representation, Elsevier B.V.; 2008. p. 88-134.
- [3] Beame P, Kautz H, Sabharwal A. Towards understanding and harnessing the potential of clause learning. J. Artif. Intell. Res. 22(2004), p. 319-351.
- [4] Kaporis Alexis C, Kirousis Lefteris M, Laïas Efthimios G. The probabilistic analysis of a greedy satisfiability algorithm. Random Struct. Algor.; 2006. 28(40) p.444-480.
- [5] Mohammad TH, Gregory S. The satisfiability threshold of random 3-SAT is at least 3.52. 2003/10/13. arXiv preprint math/0310193.
- [6] Crawford J, Auton I. Experimental results on the crossover point in satisfiability problems. Proceeding of AAAI-93, Washington, DC. 1993. p. 21-27.
- [7] Uvarov SI. An Improved generator for 3-CNF formulas. Automat. Rem. Contr.; 2020, 81(1), p. 95-103.

# Intuitive Contrasting Map for Antonym Embeddings

Igor SAMENKO<sup>a</sup>, Alexey TIKHONOV<sup>b</sup> and Ivan P. YAMSHCHIKOV<sup>c,1</sup>

<sup>a</sup>*Institute of Computational Technologies, Russian Academy of Sciences, Novosibirsk,  
Russia*

<sup>b</sup>*Yandex, Berlin, Germany*

<sup>c</sup>*LEYA Lab, Yandex and Higher School of Economics in St. Petersburg, Russia*

**Abstract.** This paper shows that modern word embeddings contain information that distinguishes synonyms and antonyms despite small cosine similarities between corresponding vectors. This information is implicitly encoded in the geometry of the embeddings and could be extracted with a straightforward manifold learning procedure or a *contrasting map*. Such a map is trained on a small labeled subset of the data and can produce new embeddings that explicitly highlight specific semantic attributes of the word. The new embeddings produced by the map are shown to improve the performance on downstream tasks.

**Keywords.** word embeddings, antonyms, manifold learning, contrasting map, word representations

## 1. Introduction

Modern word embeddings, such as [1], [2] or [3] are based on the so-called distributional hypothesis [4]. If two words are often used in a similar context, they should have a small cosine similarity between the embeddings. Naturally, such methods often fail to recognize antonyms since antonymous words, e.g., "fast" and "slow", occur in similar contexts. Many researchers address this issue from different angles.

Some authors address representations of antonyms, injecting additional information, and improving training procedures. In [5] the authors use deep learning combined with various types of semantic knowledge to produce new word embeddings that show better performance on a word similarity task. In [6] information from thesauri is combined with distributional information from large-scale unlabelled text data and obtained embeddings are used to distinguish antonyms. The authors of [7] represent semantic knowledge extracted from thesauri as many ordinal ranking inequalities and formulate the learning of semantic word embeddings as a constrained optimization problem. In [8] the authors develop these ideas further and adjust word vectors using the semantic intensity information alongside with thesauri. In [9] thesauri along with the sentiment are used to build new embeddings that contrast antonyms. In [10] authors improve the weights of feature

---

<sup>1</sup>Corresponding Author: LEYA Lab, Yandex and Higher School of Economics in St. Petersburg; Kantemirovskaya st. 3, Saint-Petersburg, Russia; E-mail: ivan@yamshchikov.info

vectors with a special method based on local mutual information and propose an extension of the skip-gram model that integrates the new vector representations into the objective function. In [11] and [12] it is shown that translation-based embeddings perform better in applications that require concepts to be organized according to similarity and better capture their true ontologic status. The authors of [13] use these ideas and demonstrate that adding a multilingual context when learning embeddings allows improving their quality via deep canonical correlation analysis.

Other researchers try to develop novel approaches that are not heavily relying on the distributional hypothesis. For example, in [14] authors introduce word-level vector representation based on symmetric patterns and report that such representations allow controlling the model judgment of antonym pairs. A special *contrasting embedding framework* is developed in [15]. While in [16] the authors train a neural network model that exploits lexico-syntactic patterns from syntactic parse trees to distinguish antonyms.

All works mentioned above were based on the assumption that antonym distinguishing information is not captured by modern word embeddings. However, this assumption is frequently questioned in the last several years. In particular, one can inject information on hyponyms, hyperonyms, synonyms, and antonyms to distinguish the obtained embeddings using additional linguistic constraints, see [17], [18] and [19]. Moreover, in [20] the authors come up with a two-phase training of a siamese network that transforms initial embeddings into the ones that clearly distinguish antonyms. While the authors of [21] develop an architecture of a distiller that extracts information on antonyms out of the pre-trained vectors.

In this work, we demonstrate that Word2Vec [1], GloVe [2], and especially FastText [3] embeddings contain information that allows distinguishing antonyms to certain extent. This information is encoded in the geometry of the obtained vector space. We propose a very simple and straightforward approach for the extraction of this information. Similarly to [20] it is based on a siamese network, yet does not require a two-phase training and is more intuitive than the one proposed in [21]. We also show that this approach could be used further to extract other semantic aspects of words out of the obtained embedding space with ease.

The contribution of this paper is as follows:

- we demonstrate that modern word embeddings contain information that allows distinguishing synonyms and antonyms;
- we show that this information could be retrieved by learning a nonlinear manifold via supervision provided by a small labeled sub-sample of synonyms and antonyms;
- we demonstrate that concatenation of these new embeddings with original embeddings improves the performance on the downstream tasks that are sensitive to synonym-antonym distinction.

## 2. Data

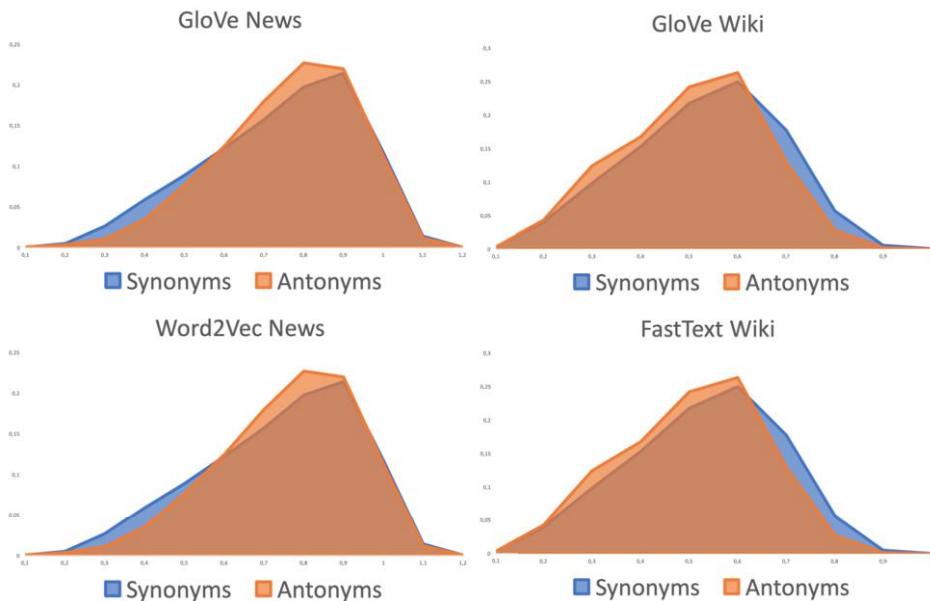
For the experiments, we used the small supervised set of synonyms and antonyms of English language provided by WordNet<sup>2</sup> that we enriched with additional data from [16]

---

<sup>2</sup><https://wordnet.princeton.edu/>

and several other publicly available sources<sup>3</sup>. We tested the methodology described below across multiple modern word embeddings, namely, FastText<sup>4</sup>, GloVe pre-trained on Wikipedia<sup>5</sup> and GloVe pre-trained on Google News<sup>6</sup> alongside with Word2Vec pre-trained on Google News<sup>7</sup>. In Figure 1 one could see initial distributions of cosines between synonyms and antonyms in four different training datasets respectively.

The WordNet dataset of synonyms and antonyms consists of 99833 word pairs. Synonymic relations are neither commutative nor transitive. For example, "economical" could be labeled as a synonym to "cheap," yet the opposite is not true<sup>8</sup>. At the same time, if "neat" is denoted as a synonym to "cool" and "cool" is denoted as a synonym to "serene," this does not imply that "neat" and "serene" are synonyms as well. All data sources used in this paper are in the public domain. To facilitate reproducibility, we share the code of the experiments<sup>9</sup>



**Figure 1.** Distribution of cosine distances between synonyms and antonyms across four different sets of embeddings.

We propose the following train-test split procedure that guarantees that the words from the training dataset do not infiltrate the test set. We add pairs to train and test with

<sup>3</sup><https://github.com/ec2604/Antonym-Detection>

<sup>4</sup><https://fasttext.cc/>

<sup>5</sup><https://nlp.stanford.edu/projects/glove/>

<sup>6</sup><https://www.kaggle.com/pkugoodspeed/nlpword2vecembeddingspretrained>

<sup>7</sup><https://github.com/mmholtz/word2vec-GoogleNews-vectors>

<sup>8</sup><https://www.thesaurus.com/browse/cheap>

<sup>9</sup><https://github.com/i-samenko/Triplet-net/>

relative frequencies of 3 to 1. If one of the words in the pair was already in the train or test, we were adding the new pair to the corresponding subset. If one word in the pair occurs both in train and in the test, we deleted such a pair. After such a test-train split, we obtained 80 080 pairs. 65 292 pairs of 26 264 unique words formed the training dataset, and 14 788 pairs of 8737 unique words comprised the test dataset.

Figure 1 seems to back up the widespread intuition that modern embedding can not distinguish synonyms and antonyms. However, in the next sections, this paper demonstrates that this statement does not hold.

### 3. Learning Contrasting Map

If one assumes that information allowing to distinguish synonyms and antonyms is already present in the raw embeddings, one could try to extract it by building a manifold learning procedure that would take original embedding as input and try to map it in a new space of representations, where the synonym-antonym contrast becomes explicit.

The initial embedding space is  $\mathbb{R}^m$  with a distance  $D_m$  defined on it, and for every word ' $w$ ', for any of its synonyms ' $s$ ', and for any of its antonyms ' $a$ ' the following holds  $D_m(w, s) \simeq D_m(w, a)$ . A new embedding space of lower dimension  $\mathbb{R}^k$  has a corresponding distance  $D_k$ . One would like to find a map  $f : \mathbb{R}^m \rightarrow \mathbb{R}^k$  such that the following holds  $D_k(w, s) < D_k(w, a)$  in a new  $\mathbb{R}^k$  embedding space.

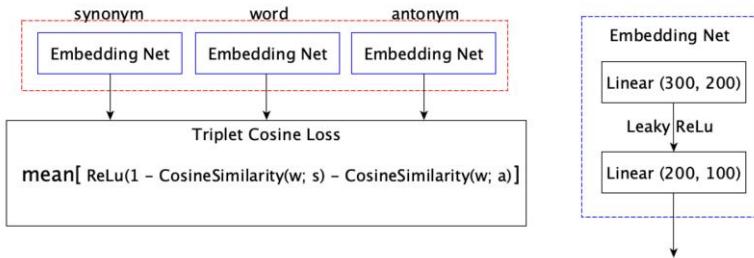
$$f = \begin{cases} f : \mathbb{R}^m \rightarrow \mathbb{R}^k, & m \gg k; \\ D_k(w, s) < D_k(w, a), & \forall w, s, a. \end{cases} \quad (1)$$

Since the amount of synonyms and antonyms in any given language is growing excessively with the growth of the training sample of texts, one can not check these conditions for every word pair explicitly. One can only use a labeled subset of the vocabulary, where synonyms and antonyms are contrasted already, so it is hard to establish a procedure that would guarantee Inequalities 1, hence we use  $\lesssim$  for the conditions. At the same time, despite the limited size of the training dataset, one would hope that the obtained representations are general enough to distinguish the synonyms and antonyms that are not included in the training data.

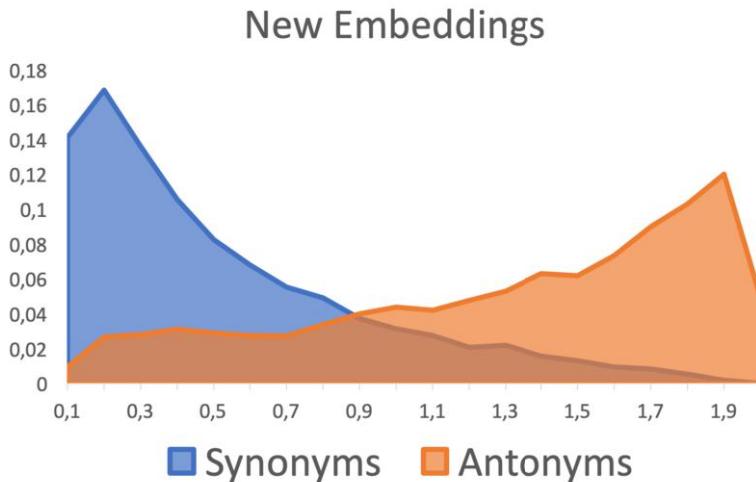
To train such a map let us regard an architecture, shown in Figure 2. It is a 'Siamese' network [22] where weights are shared across three identical EmbeddingNets. Each EmbeddingNet maps the word ' $w$ ', its synonym ' $s$ ' and its antonym ' $a$ ' respectively. The resulting cosine similarities between synonyms and antonyms are simply included in the loss function in such a way that  $D_k(w, s)$  is minimized and  $D_k(w, a)$  is maximized explicitly. The whole system is trained end-to-end on 65 292 pairs of synonyms and antonyms described in Section 2.

### 4. Experiments

First of all, let us check if the condition listed in Equation 1 is satisfied in the transformed embedding space  $\mathbb{R}^k$ . Figure 3 illustrates the distributions of the cosine distances



**Figure 2.** Siamese Triplet Network trained to distinguish synonyms and antonyms. EmbeddingNet is the contrasting map  $f : \mathbb{R}^m \rightarrow \mathbb{R}^k$ . The weights of three EmbeddingNets are shared in the end-to-end training. The resulting architecture is trained to minimize cosine similarities between synonyms and maximize the cosine similarities between antonyms in the transformed low-dimensional embeddings space  $\mathbb{R}^k$ .



**Figure 3.** Distribution of cosine distances between synonyms and antonyms in the transformed space  $\mathbb{R}^k$  for FastText. Test set. Different datasets produce similar results. Distances between synonyms tend to become smaller, distances between antonyms tend to increase.

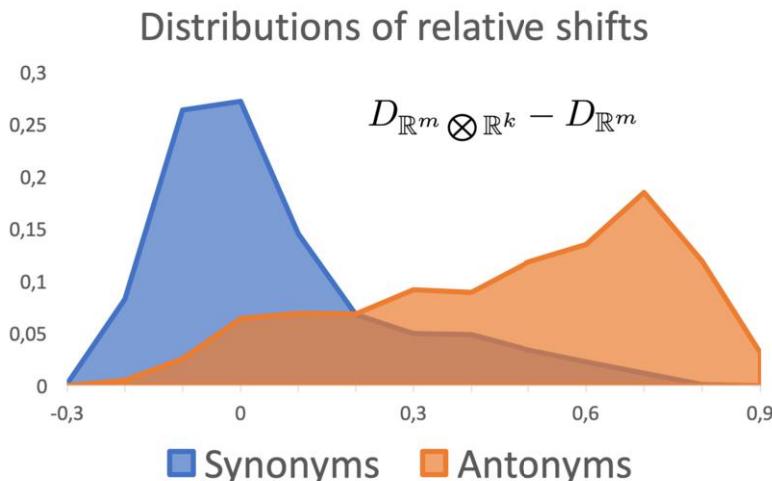
between synonyms and antonyms in  $\mathbb{R}^k$  for English FastText embeddings. The situation is drastically improved in contrast with raw embeddings shown in Figure 1.

One can have a close look at the tails of the distributions shown in Figure 3. To simplify further experiments and improve reproducibility we also publish the resulting distances for the test set<sup>10</sup>.

<sup>10</sup><https://github.com/i-samenko/Triplet-net/>

Here are some examples of word pairs that were marked as antonyms in the test dataset, yet are mapped close to each other by the contrast map: sonic — supersonic; fore — aft; actinomorphic — zygomorphic; cable — hawser; receive — give; ceiling — floor. Here are some examples of word pairs that were marked as synonyms in the test dataset, yet are mapped far of each other by the contrast map: financial — fiscal; mother — father; easy — promiscuous; empowered — sceptred; formative — plastic; frank — wiener; viii — eighter; wakefulness — sleeplessness. One can see that some of the contrasting map errors are due to the debatable labeling of the test dataset, others occur with the words that are rare.

To be sure that other properties of the original embeddings are preserved we concatenate new embedding with the old, raw ones. Figure 4 depicts the difference of the pairwise distance between synonyms and antonyms in the space of concatenated embeddings  $D_{\mathbb{R}^m \oplus \mathbb{R}^k}$  and in the space of raw embeddings  $D_{\mathbb{R}^m}$ . The distributions are obtained for the test subset of data. The map did not see these word pairs in training.



**Figure 4.** Cosine distances between synonyms and antonyms in the raw embeddings and in the space where they are concatenated with the new ones. FastText embeddings. Test set.

We train an XGBoost classifier on four different raw embeddings and check the resulting accuracy of the classifiers on the test subset of synonym and antonym pairs. Table 1 clearly shows that the accuracy of a classifier trained on raw embeddings is consistently lower than the accuracy of the same classifier trained on the newly transformed embeddings, produced by the EmbeddingNet. One can also see that a classifier trained on the concatenation of the raw embeddings with the new ones also outperforms the classifier trained solely on the original embeddings.

FastText embeddings are capturing more than 80% of synonym-antonym relations with the proposed contrasting map, and more than 70% out of these relations are captured out of the box. GloVe embeddings seem to contain the least information on the synonym-

**Table 1.** Comparison of four different embeddings. For every type of embedding, XGBoost classifier is trained to distinguish two input vectors as synonyms or antonyms.

Embeddings type	Raw	New	Concatenated
Word2Vec	0.67	0.85	0.81
GloVe Wiki	0.65	0.75	0.72
GloVe Google News	0.67	0.84	0.78
FastText	0.73	<b>0.88</b>	<b>0.85</b>

**Table 2.** Concatenation of the original FastText embeddings with transformed embeddings improves the accuracy of logistic regression-based classifiers trained on various datasets.

Dataset	FastText only	Concatenated
IMDB reviews	0.86	0.88 ( <b>+2.2%</b> )
Cornell reviews	0.75	0.76 ( <b>+1.0%</b> )
Toxic Comments	0.94	0.95 ( <b>+0.6%</b> )
MDSD books	0.69	0.77 ( <b>+11.3%</b> )
MDSD DVDs	0.70	0.76 ( <b>+8.0%</b> )
MDSD electronic	0.72	0.78 ( <b>+9.4%</b> )
MDSD kitchen	0.78	0.80 ( <b>+3.4%</b> )
MDSD all categories	0.76	0.79 ( <b>+3.6%</b> )

antonym relations. Further experiments are conducted on FastText embeddings since they capture the most out of synonym-antonym relations.

To illustrate the potential usage of such embeddings obtained with a contrasting map we run a series of experiments with various NLP datasets that intuitively might need to contrast synonyms and antonyms for the successful performance: binary sentiment classifier for IMDB reviews<sup>11</sup>, binary sentiment classifier for Cornell movie reviews<sup>12</sup>, binary classifier to identify toxic comments<sup>13</sup>, sentiment classifiers across several thematic domains of Multi-Domain Sentiment Dataset<sup>14</sup>.

For every dataset, we trained a logistic regression using pre-trained FastText embeddings and measured its accuracy on the test. Then we retrained the same logistic regression with new concatenated embeddings. Table 2 demonstrates how the usage of the transformed embeddings improves the accuracy on various datasets.

## 5. Discussion

The proposed methodology demonstrates that contrary to common intuition modern word embeddings contain information that allows distinguishing synonyms and antonyms. The approach could possibly be scaled to other semantic aspects of the words. In its most general form, the approach allows mapping original embeddings into spaces of lower dimensions that could explicitly highlight certain semantic aspects using a la-

<sup>11</sup><https://ai.stanford.edu/~amaas/data/sentiment/>

<sup>12</sup><http://www.cs.cornell.edu/people/pabo/movie-review-data/>

<sup>13</sup><https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge>

<sup>14</sup><http://www.cs.jhu.edu/~ndredze/datasets/sentiment/>

beled dataset of a limited size. This semantic information can be effectively incorporated into the downstream tasks.

Conceptually, the proposed methodology allows for revisiting the questions of language acquisition in the context of the distributional hypothesis. If one assumes that semantic information attached to a given word is not a rigid structure but depends on the training corpus, it seems that modern embeddings capture these diverse semantic fields successfully, provided the corpus is large enough. This result does not mean that such semantic aspects are explicit and could be immediately extracted out of the embeddings. The spaces of modern word embeddings could be profoundly nonlinear concerning a given semantic attribute of the word. A deeper understanding of the geometric properties of these spaces could significantly improve the quality of the resulting models. Indeed, the very assumption that semantic similarity could be captured with cosine distance in Euclidian space is debatable.

Though the choice of the embedding space and the notion of distance on it both need further, more in-depth investigations, this paper demonstrates the simple methods of representational learning applied to the raw embeddings can distill this implicitly encoded information reasonably well.

## 6. Conclusion

This paper demonstrates that, contrary to a widely spread opinion, modern word embeddings contain information that allows distinguishing synonyms from antonyms. This information is encoded in the geometry of the embeddings and could be extracted with manifold learning. The paper proposes a simple and intuitive approach that allows obtaining a *contrasting map*. Such a map could be trained on a small subset of the vocabulary and is shown to highlight relevant semantic information in the resulting vector embedding. The new embeddings, in which the information on synonyms and antonyms is disentangled, improve the performance on the downstream tasks. The proposed methodology of contrasting maps could potentially be further extended to other semantic aspects of the words.

## References

- [1] Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:13013781. 2013.
- [2] Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP); 2014. p. 1532–1543.
- [3] Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics. 2017;5:135–146.
- [4] Harris Z. Distributional Hypothesis. Word. 1954;10(23):146–162.
- [5] Bian J, Gao B, Liu TY. Knowledge-powered deep learning for word embedding. In: Joint European conference on machine learning and knowledge discovery in databases. Springer; 2014. p. 132–148.
- [6] Ono M, Miwa M, Sasaki Y. Word embedding-based antonym detection using thesauri and distributional information. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; 2015. p. 984–989.
- [7] Liu Q, Jiang H, Wei S, Ling ZH, Hu Y. Learning semantic word embeddings based on ordinal knowledge constraints. In: Proceedings of the 53rd Annual Meeting of the Association for Computational

- Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers); 2015. p. 1501–1511.
- [8] Kim JK, de Marneffe MC, Fosler-Lussier E. Adjusting word embeddings with semantic intensity orders. In: Proceedings of the 1st Workshop on Representation Learning for NLP; 2016. p. 62–69.
  - [9] Dou Z, Wei W, Wan X. Improving word embeddings for antonym detection using thesauri and Senti-WordNet. In: CCF International Conference on Natural Language Processing and Chinese Computing. Springer; 2018. p. 67–79.
  - [10] Nguyen KA, im Walde SS, Vu NT. Integrating Distributional Lexical Contrast into Word Embeddings for Antonym-Synonym Distinction. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers); 2016. p. 454–459.
  - [11] Hill F, Cho K, Jean S, Devin C, Bengio Y. Not all neural embeddings are born equal. arXiv preprint arXiv:14100718. 2014.
  - [12] Hill F, Cho K, Jean S, Devin C, Bengio Y. Embedding word similarity with neural machine translation. arXiv preprint arXiv:14126448. 2014.
  - [13] Lu A, Wang W, Bansal M, Gimpel K, Livescu K. Deep multilingual correlation for improved word embeddings. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; 2015. p. 250–256.
  - [14] Schwartz R, Reichart R, Rappoport A. Symmetric pattern based word embeddings for improved word similarity prediction. In: Proceedings of the nineteenth conference on computational natural language learning; 2015. p. 258–267.
  - [15] Chen Z, Lin W, Chen Q, Chen X, Wei S, Jiang H, et al. Revisiting word embedding for contrasting meaning. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers); 2015. p. 106–115.
  - [16] Nguyen KA, Schulte im Walde S, Vu NT. Distinguishing Antonyms and Synonyms in a Pattern-based Neural Network. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Valencia, Spain; 2017. .
  - [17] Mrkšić N, Vulić I, Séaghdha DÓ, Leviant I, Reichart R, Gašić M, et al. Semantic specialization of distributional word vector spaces using monolingual and cross-lingual constraints. Transactions of the association for Computational Linguistics. 2017;5:309–324.
  - [18] Vulić I, Mrkšić N. Specialising Word Vectors for Lexical Entailment. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers); 2018. p. 1134–1145.
  - [19] Vulić I. Injecting lexical contrast into word vectors by guiding vector space specialisation. In: Proceedings of The Third Workshop on Representation Learning for NLP; 2018. p. 137–143.
  - [20] Etcheverry M, Wonsever D. Unraveling Antonym’s Word Vectors through a Siamese-like Network. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics; 2019. p. 3297–3307.
  - [21] Ali MA, Sun Y, Zhou X, Wang W, Zhao X. Antonym-synonym classification based on new sub-space embeddings. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33; 2019. p. 6204–6211.
  - [22] Bromley J, Guyon I, LeCun Y, Säckinger E, Shah R. Signature verification using a “siamese” time delay neural network. In: Advances in neural information processing systems; 1994. p. 737–744.

# DYPLODOC: Dynamic Plots for Document Classification

Anastasia MALYSHEVA<sup>a</sup>, Alexey TIKHONOV<sup>b</sup> and Ivan P. YAMSHCHIKOV<sup>c,1</sup>

<sup>a</sup>*Open Data Science, Moscow, Russia*

<sup>b</sup>*Yandex, Berlin, Germany*

<sup>c</sup>*LEYA Lab, Yandex and Higher School of Economics in St. Petersburg, Russia*

**Abstract.** Narrative generation and analysis are still on the fringe of modern natural language processing yet are crucial in a variety of applications. This paper proposes a feature extraction method for plot dynamics. We present a dataset that consists of the plot descriptions for thirteen thousand TV shows alongside meta-information on their genres and dynamic plots extracted from them. We validate the proposed tool for plot dynamics extraction and discuss possible applications of this method to the tasks of narrative analysis and generation.

**Keywords.** narrative analysis, plot dynamics, plot structure, plot extraction

## 1. Introduction

Natural Language Generation (NLG) is one of the areas of Natural Language Processing (NLP). Deep learning enabled various generative applications where generated texts are short and constrained by the context. These could be dialogue responses [1,2], traditional or stylized poetry [3,4,5], end-to-end author or source stylization [6,7,8]. One of the bottle-necks that holds modern Natural Language Processing (NLP) from the generation of longer texts is the concept of *narrative*. There are constant attempts to generate longer blocks of text, such as [9] or [10], yet they succeed under certain stylistic and topical constraints that exclude the problem of narrative generation altogether [11].

Though philosophers and linguists try to conceptualize the notions of plot, narrative arc, action, and actor for almost a century [12,13,14], few of these theoretical concepts could be instrumental for modern NLP. [15] present a machine comprehension corpus for the end-to-end evaluation of script knowledge. The authors demonstrate that though the task is not challenging to humans, existing machine comprehension models fail to perform well, even if they make use of a commonsense knowledge base. Despite these discouraging results, there are various attempts to advance narrative generation within the NLP community, see [16,17]. For a detailed review of approaches to narrative generation, we address the reader to [18], here we just mention several ideas that are relevant within the scope of this paper.

---

<sup>1</sup>Corresponding Author: LEYA Lab, Yandex and Higher School of Economics in St. Petersburg; Kantemirovskaya st. 3, Saint-Petersburg, Russia; E-mail: ivan@yamshchikov.info

One line of thought is centered around some form of a multi-agent system with various constraints that keep generated texts within some proximity of human-written narrative. For example, the paper [19] presents a multi-agent framework for automatic story generation where plots are automatically created by semi-autonomous characters while the resulting plot is converted into text by a narrator agent. [20] experiment with a dialogic form of a story: a deep representation of a story is converted into a dialog that unfolds between two agents with distinctive stylistic characteristics. Such multiagent-based systems show a lot of potential yet do not leverage the existing datasets that include an extensive amount of narratives written by humans.

Another line of research tries to use some NLP methods and statistical procedures to systematize existing narratives using some form of synthetic features that could in some form describe the narrative structure of a given story. [21] represents a story as a cluster of emotional links and tensions between characters that progress over story-time. [22] introduce a concept of narrative arc in a context of large-scale text mining. The author conceptualizes plot structure as a path through a multidimensional space derived from a topic model. This approach allows us to see plot dynamics within any given book or show in detail, yet it is harder to generalize. Topic modeling tends to provide profoundly different sets of topics across genres. When aggregating across all topics within a dataset one gets a high dimensional space that is harder to interpret. For example, similar narrative arcs of a detective story would look different in terms of topic modeling if the same plot takes place in Victorian England or on Mars. One would like to keep the concept of a narrative arc yet to make it more abstract in order to be able to carry comparative narrative research across genres and settings. [23] develop ideas that address this challenge to a certain extent. The authors mathematize a concept of narrative arc that depicts the evolution of an agent's belief over a set of the so-called *universes*. The authors estimate these probabilities with pre-trained probabilistic topic classifiers and combine these estimates with a multi-agent dialogic approach. They demonstrate that the agent that tracks the dynamics of such narrative arcs generates more engaging dialog responses.

This paper shows that the concept of the narrative arc could be applied to a vast variety of narratives in natural languages and does not have to be limited to the setting of a dialogue system. The contribution of this paper is two-fold: (1) we provide a large dataset of longer storylines that could be used for narrative research and (2) use it to demonstrate how the notion of the narrative arc could be applied to longer texts and provide meaningful insights into the structure of the story. This structural insight could be useful both for quantitative research of narrative dynamics and as synthetic features for the generation of narrative structures in the future.

## 2. Data

TVmaze<sup>2</sup> is a huge database with the data on different TV shows that include the storyline as well as certain meta information. Using TVmaze API we have obtained 300+ Mb of data that could be found online as a part of the DYPLODOC package<sup>3</sup>. The dataset includes 13 813 TV shows and 21 962 seasons that consist of more than 300 000 episodes.

---

<sup>2</sup><https://www.tvmaze.com/>

<sup>3</sup><https://github.com/AnastasiaMalysheva/dyplodoc>

Every episode in the dataset is annotated with a condensed description of the main storyline and the narrative structure of a season is defined as a sequence of such descriptions. The data includes main characters, that could be extracted with the NER algorithm, the majority of texts are centered around most reportable events [24], and the narrative is split into clear steps, certain 'quanta' of a story. Almost every season in TVmaze has a tag of a genre, such as (Horror, Food, Drama, etc). We hope that the publication of the dataset could facilitate further quantitative research of narrative structures.

### 3. Plot Dynamics Extraction Tool

Here we leverage the fact that episode descriptions in the TVmaze dataset form seasons and demonstrate how one could build the narrative arc of a season by constructing artificial features that would to a certain degree describe the season's plot dynamics. We call this method Plot Dynamics Extraction (PDE) Tool.

Let us regard a specific TV show  $X$  as a set of  $N$  seasons  $X : \{X_1, X_2, \dots, X_N\}$ . Every season  $X_i$  could be represented as a sequence of episodes  $X_i : \{y_1, y_2, \dots, y_M\}$ . Every  $y_i$  is a description of one particular episode.

We investigated the plot dynamics using two different pre-trained probabilistic topic classifiers. However, unlike [22] we choose fixed topics that form a universal state space of narrative arcs. One could discuss if the provided list of dimensions is complete, yet we believe that two classifiers are sufficient to extract non-trivial narrative arcs out of a variety of texts:

- Classifier of movie genres trained on Wikipedia film synopses provided in [25]. The labels of genres include: Comedy, Drama, Western, Adventure, Animation, Action, Thriller, Family, Romance, Fantasy, Horror, History, Music, Sci-Fi, War, Crime, Musical, Biography, Mystery, Sport.
- DeepMoji: Deep neural network that was trained on a corpus of text that included 64 emojis, see [26]. In their blogpost<sup>4</sup> the authors suggest hierachic aggregation of 64 emojis. Using this aggregation we limit ourselves to 11 labels in this work. These labels include: Love, Happy, Wink, Deal, Force, Eyes, Fear, Mad, Sad, Music, Misc.

Thus every narrative arc is a trajectory on a union of two simplexes: a 20-simplex of genres and an 11-simplex of emotions. For the vast majority of the cases, the classifier mostly detects one or two most dominant genres and topics respectively, limiting the meaningful development of a given narrative art to one plane on each simplex.

Similarly to [23] we propose to regard a finite set of universes,  $\mathcal{U}$ . One could develop a universe model to assess the compatibility of an episode  $y_i$  with a given universe,  $u \in \mathcal{U}$ . Given such a model, one could update the posterior universe distribution over a sequence of episodes  $y_i$ . For each universe  $u$ , the universe model assigns a likelihood  $p(y_i|y_{1:i-1}, u)$  to an episode  $y_i$ , conditioned on previous episodes  $y_{1:i-1}$  of the season. Introducing  $z(u)$  as the prior distribution over universes, the conditional probability  $z(u|y_i)$  could be defined as follows  $z(u|y_i) = z(u) \times \frac{z(y_i|u)}{z(y_i)}$ .

Using this notation [23] obtain the following approximation

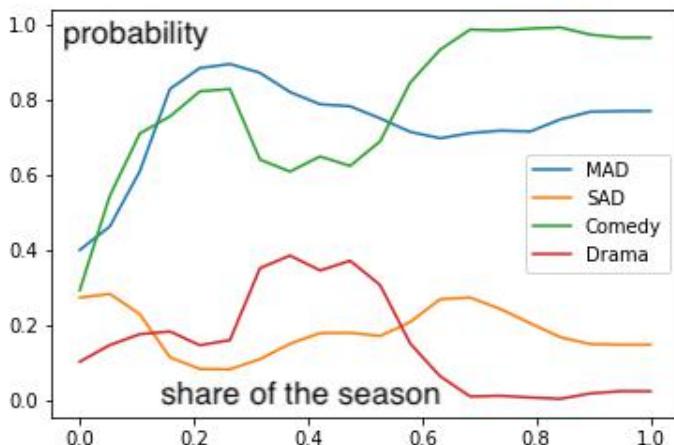
---

<sup>4</sup><https://medium.com/@bjarkefelbo/what-can-we-learn-from-emojis-6beb165a5ea0>

$$p(u|y_{1:i}) = p(u|y_{1:i-1}) \times \frac{z(u|y_i)}{\sum_{u'} p_{i-1}(u'|y_{1:i-1})z(u'|y_i)} \quad (1)$$

In order to guarantee that the sequence of  $p(u|y_{1:i})$  forms a series of coordinates on the union of 20-simplex of genres and 11-simplex of emoji, we introduce two additional smoothing procedures. First, we control  $z(u)$  distribution update so that it does not degenerate into zero, since in this case Equation (1) becomes obsolete. We also use exponential smoothing average instead of standard multiplicative update.

Figure 1 illustrates the resulting dynamics of one season of 'BoJack Horseman', extracted with a Deepemoji classifier and with a classifier pre-trained on the genres<sup>5</sup>. There are four major non-zero components. On the emoji-simplex the narrative is balancing between MAD and SAD, though the former is prevailing throughout the season, the latter peaks in the episode when BoJack's daughter overdoses and is found in the hospital. On the genre-simplex, the narrative is balancing between Comedy and Drama. BoJack Horsemen is a dark comedy animation series so it stands to reason that Comedy dominates the season while Drama peaks in the second quarter of the season when BoJack spends time with his elderly mother in the nursing home.



**Figure 1.** Dynamics of the fourth season of 'BoJack Horseman'. Tags from the simplex of emotions are denoted with capital letters.

#### 4. Experimenting with PDE Tool

Though scores of pre-trained deepemoji and genre classifiers have very intuitive interpretation it is hard to claim that they could be useful for any form of comparative narrative

<sup>5</sup>PDE Tool is available as a part of DYPLODOC package <https://github.com/AnastasiaMalysheva/dyplodoc>

research. The dynamic aspect of these constructed features is clear yet it is not obvious that their dynamics are meaningful in any non-trivial sense. We compare dynamic plots extracted from the season description with static meta-information that is included in the TVmaze data<sup>6</sup>. If different episodes of the season are labeled with a different tag we choose the most frequent tag as the tag for the whole season. If the show was tagged with two values<sup>7</sup>, we treated it like one static tag. For example, 'BoJack Horseman' has Drama | Comedy as a genre, thus Drama, Comedy and Drama | Comedy for three separate static tags.

PDE-tool represents every plot as an array of tags (such as genres or emoticons), in which every tag is represented as a sequence of probability estimations ranging from 0 to 1. We concatenate these representations with extra synthetic features, namely: mean, median, minimum, maximum, and variance for the probability of every tag; area under the curve; the moment of intersection with another tag probability estimation, etc. Thus we represent every text as a multidimensional vector. Using these dynamic based features we clustered the seasons, applying standard K-means clustering over a PCA<sup>8</sup>. We have carried our experiments across all possible combinations of features, a different number of principal components (50, 100, 200, 300), and a various number of clusters (50, 75, 100, 150, 200, 250, 300). Then we calculated mutual information between obtained clusters and static tags for the corresponding seasons. The higher values of mutual information were obtained for the following setup: synthetic features included statistics and area under the curve for every component; the number of principal components equals 200; the number of clusters for K-means equals 200.

With this setup, the mutual information between clusters based on both classifier combined and static tags was 0.57, as shown in Table 1. One could see that both the dynamics based on the genre classifier and the emoji classifier-based dynamics allow inferring static genre tags. The combination of two classifiers yields higher mutual information. Dynamic plots extracted by different classifiers differ significantly and go beyond static genre tags.

## 5. Discussion

Figure 2 shows the resulting t-SNE (t-distributed stochastic neighbor embedding) visualization, see [27] for details, of the clustering based on plot dynamics. The centers of the clusters are labeled with static genre tags. The average size of a cluster in the resulting clustering is 109 and the average share of the dominant tag is 65%. The smallest of the clusters consists of 9 show seasons, the biggest includes 288 seasons. In some clusters the share of the dominant tag reaches 100%, the lowest possible percentage of the dominant tag is 42%.

Using synthetic features based on the PDE tool one could partially infer static human labels for season's genre. Table 1 demonstrates that the dynamic plot structure is non-

<sup>6</sup>7911 seasons out of 21962 have no tags so we exclude them from this experiment.

<sup>7</sup>We filtered the shows that had more than 2 tags for their genre.

<sup>8</sup>Since the length of the seasons might differ and we are interested in comparative research of the obtained arcs we scale every series to the length of ten episodes either uniformly stretching or compressing the resulting timeline.

**Table 1.** Mutual Information (denoted as MI) between k-means clusters based on the features obtained with plot dynamics extraction and the clusters defined by static genre meta-information alongside with the silhouette for the obtained clusterizations.

	MI	Silhouette
Dynamic Genres	0.47	0.23
Dynamic Emojis	0.36	0.16
Combined		
Dynamic Features	0.57	0.20

**Table 2.** Three biggest clusters in the resulting clusterization that have the highest percentage of a dominant tag.

# of shows	Dominant Static Tag	% of Dominant Static Tag
274	Comedy   Drama"	74
260	Drama   Fantasy"	79
253	Drama   Comedy"	72

trivial and could be useful for quantitative research of narrative. We publish extracted plot dynamics along with the TVmaze data to facilitate further research.

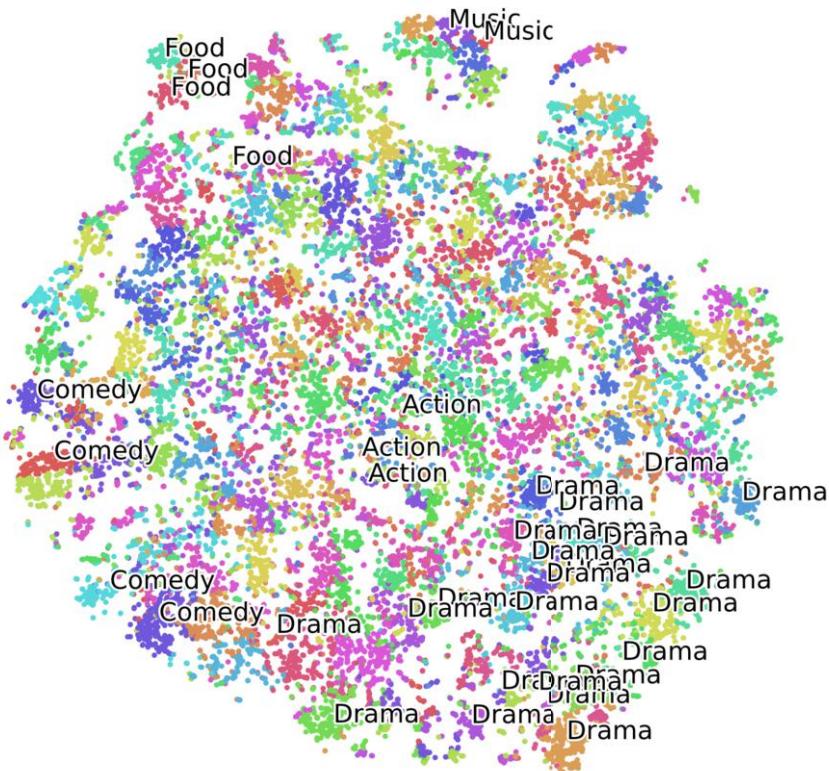
Now let us show how the PDE tool could be used for comparative research of narrative. Let us zoom in on the 200 clusters obtained from dynamics plots extracted from the descriptions of the seasons. Table 2 shows the three biggest resulting clusters with the highest percentage of one static genre tags.

The first cluster could be attributed to darker comedies. Figure 1 shown above is a good representative of this cluster. The second cluster could be attributed to historic dramas. The cluster is characterized by the dominance of Drama and interplay of Sad and Mad emojis. Figure 3 illustrates it with the dynamics of the second season of Victoria. The third cluster is also dominated by Drama but with the certain presence of Comedy as well as with domination of Mad over Sad, see Figure 4 for an example. These are only a few examples that illustrate how rich is the information provided by the PDE tool and its potential for quantitative analysis of narrative.

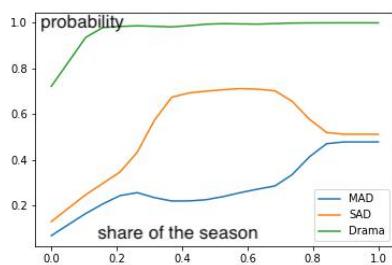
## 6. Conclusion

This work makes the following contributions: (1) a universal feature extraction method for plot dynamics; (2) a dataset of natural language plot descriptions; (3) finally, the paper shows that plot dynamics features extracted with the proposed method are meaningful and can shed light on the general structure of the narrative.

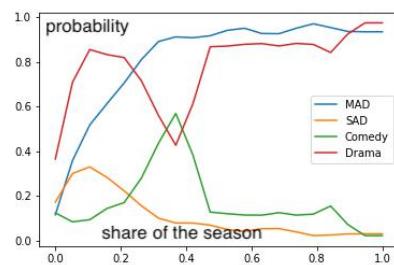
We hope that this paper could help to revive the attention of the community to the problems of narrative analysis and generation at the very least by providing data and tools for further exploration. We also hope to use narrative arch extracted with PDE tool as a form of semi-supervision for narrative-oriented generative models. The extracted dynamic plots are presented alongside the textual data. We hope that the proposed tool would inspire further research on plot dynamics and inform the discussions on narrative structure, particularly in the context of narrative generation.



**Figure 2.** t-SNE visualization of the clustering based on plot dynamics. The labels mark clusters in which the dominant genre tag corresponds to 70% or more shows in the cluster.



**Figure 3.** Dynamics of the second season of 'Victoria'. Tags from the simplex of emotions are denoted with capital letters.



**Figure 4.** Dynamics of the third season of 'Orange is a new black'. Tags from the simplex of emotions are denoted with capital letters.

## References

- [1] Li J, Monroe W, Ritter A, Jurafsky D, Galley M, Gao J. Deep Reinforcement Learning for Dialogue Generation. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing; 2016. p. 1192–1202.

- [2] Li Z, Kiseleva J, de Rijke M. Dialogue generation: From imitation learning to inverse reinforcement learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33; 2019. p. 6722–6729.
- [3] Zhang X, Lapata M. Chinese poetry generation with recurrent neural networks. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); 2014. p. 670–680.
- [4] Wang DZ, He W, Wu H, Wu H, Li W, Wang H, et al. Chinese Poetry Generation with Planning based Neural Network. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers; 2016. p. 1051–1060.
- [5] Tikhonov A, Yamshchikov IP. Guess who? Multilingual approach for the automated generation of author-stylized poetry. In: 2018 IEEE Spoken Language Technology Workshop (SLT). IEEE; 2018. p. 787–794.
- [6] Hu Z, Yang Z, Liang X, Salakhutdinov R, Xing EP. Toward Controlled Generation of Text. In: International Conference on Machine Learning; 2017. p. 1587–1596. Available from: <http://proceedings.mlr.press/v70/hu17e.html>.
- [7] Tikhonov A, Shibaev V, Nagaev A, Nugmanova A, Yamshchikov IP. Style transfer for texts: Retrain, report errors, compare with rewrites. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); 2019. p. 3927–3936.
- [8] Syed B, Verma G, Srinivasan BV, Natarajan A, Varma V. Adapting Language Models for Non-Parallel Author-Stylized Rewriting. In: AAAI; 2020. p. 9008–9015.
- [9] Kedziora RK. Understanding and Generating Multi-Sentence Texts; 2019.
- [10] Agafonova Y, Tikhonov A, Yamshchikov IP. Paranoid Transformer: Reading Narrative of Madness as Computational Approach to Creativity. Future Internet. 2020;12(11):182.
- [11] van Stegeren J, Theune M. Narrative Generation in the Wild: Methods from NaNoGenMo. In: Proceedings of the Second Workshop on Storytelling; 2019. p. 65–74.
- [12] Shklovsky V. Theory of prose (B. Sher, Trans.). Champaign, IL: Dalkey Archive Press Original work published. 1925.
- [13] Propp V. Morphology of the Folktale, trans. Louis Wagner, 2d ed. 1968.
- [14] Van Dijk TA. Philosophy of action and theory of narrative. Poetics. 1976;5(4):287–338.
- [15] Ostermann S, Roth M, Pinkal M. MCScript2. 0: A Machine Comprehension Corpus Focused on Script Events and Participants. In: Proceedings of the Eighth Joint Conference on Lexical and Computational Semantics (\* SEM 2019); 2019. p. 103–117.
- [16] Fan A, Lewis M, Dauphin Y. Strategies for Structuring Story Generation. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics; 2019. p. 2650–2660.
- [17] Ammanabrolu P, Tien E, Cheung W, Luo Z, Ma W, Martin LJ, et al. Story realization: Expanding plot events into sentences. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34; 2020. p. 7375–7382.
- [18] Kybartas B, Bidarra R. A survey on story generation techniques for authoring computational narratives. IEEE Transactions on Computational Intelligence and AI in Games. 2016;9(3):239–253.
- [19] Theune M, Rensen S, op den Akker R, Heylen D, Nijholt A. Emotional characters for automatic plot creation. In: International Conference on Technologies for Interactive Digital Storytelling and Entertainment. Springer; 2004. p. 95–100.
- [20] Walker MA. M2D: Monolog to Dialog Generation for Conversational Story Telling. In: Interactive Storytelling: 9th International Conference on Interactive Digital Storytelling, ICIDS 2016, Los Angeles, CA, USA, November 15–18, 2016, Proceedings. vol. 10045. Springer; 2016. p. 12.
- [21] Pérez y Pérez R. Employing emotions to drive plot generation in a computer-based storyteller. Cognitive Systems Research. 2007;8(2):89–109.
- [22] Schmidt BM. Plot archeology: A vector-space model of narrative structure. In: 2015 IEEE International Conference on Big Data (Big Data). IEEE; 2015. p. 1667–1672.
- [23] Mathewson KW, Castro PS, Cherry C, Foster G, Bellemare MG. Shaping the Narrative Arc: An Information-Theoretic Approach to Collaborative Dialogue. arXiv preprint arXiv:190111528. 2019.
- [24] Ouyang J, McKeown K. Modeling reportable events as turning points in narrative. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing; 2015. p. 2149–2158.
- [25] Hoang Q. Predicting movie genres based on plot summaries. arXiv preprint arXiv:180104813. 2018.
- [26] Felbo B, Mislove A, Søgaard A, Rahwan I, Lehmann S. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In: Proceedings of the 2017

Conference on Empirical Methods in Natural Language Processing; 2017. p. 1615–1625.

- [27] Van der Maaten L, Hinton G. Visualizing data using t-SNE. Journal of machine learning research. 2008;9(11).

# Chained Digital Signature for the Improved Video Integrity Verification

Linju LAWRENCE<sup>a,1</sup> and R SHREELEKSHMI<sup>b</sup>

<sup>a</sup> Department of Computer Science and Engineering, College of Engineering, Trivandrum (Affiliated to APJ Abdul Kalam Technological University), Thiruvananthapuram - 695016, Kerala, India

<sup>b</sup> Department of Information Technology, Government Engineering College, Barton Hill (Affiliated to APJ Abdul Kalam Technological University), Thiruvananthapuram - 695035, Kerala, India

**Abstract.** The recorded videos from the surveillance cameras can be used as potential evidence in forensic applications. These videos can be easily manipulated or tampered with video editing tools without leaving visible clues. Hence integrity verification is essential before using the videos as evidence. Existing methods mostly depend on the analysis of video data stream and video container for tampering detection. This paper discusses an active video integrity verification method using Elliptic Curve Cryptography and blockchain. The method uses Elliptic Curve Digital Signature Algorithm for calculating digital signature for video content and previous block. The digital signature of the encoded video segment (video content with predetermined size) and that of previous block are kept in each block to form an unbreakable chain. Our method does not consider any coding or compression artifacts of the video file and can be used on any video type and is tested on public-available standard videos with varying sizes and types. The proposed integrity verification scheme has better detection capabilities towards different types of alterations like insertion, copy-paste and deletion and can detect any type of forgery. This method is faster and more resistant to brute force and collision attacks in comparison to existing recent blockchain method.

**Keywords.** Blockchain, Elliptic curve digital signature, Hash function, Video integrity.

## 1. Introduction

A video record can be used as a primary source of evidence in digital forensics. However, the problem is that it can undergo various tampering attacks. In such a scenario, identifying authentic video from the forged one is a challenging task. So, the verification of integrity of the video is crucial before it can be used as evidence. Forgery detection in digital videos is classified into intra-frame tampering detection and inter-frame tampering detection. Inter-frame tampering detection includes frame-duplication/frame-deletion/frame-insertion detection and temporal interpolation deletion. Frame-insertion/duplication/deletion detection uses sensor, recompression, motion, brightness and pixel level features. Intra-frame tampering detection is categorized as upscale crop

---

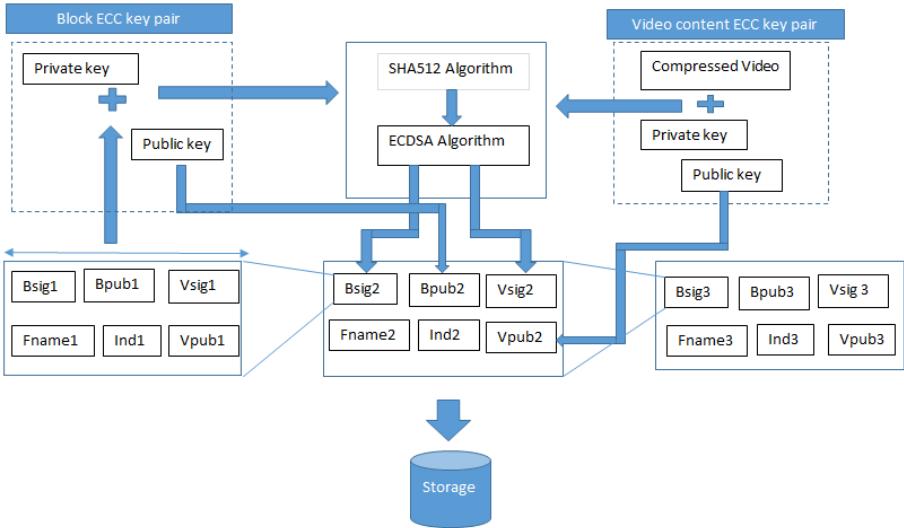
<sup>1</sup> Corresponding Author: Department of Computer Science and Engineering, College of Engineering Trivandrum, Thiruvananthapuram – 695016, Kerala, India; E-mail: linjulawrence680@gmail.com.

detection and copy-move detection. This copy-move detection method considers pixel-similarity/object/motion features. Different features used for forgery detection in digital videos are sensor/camera artifacts and coding/motion/object features [1]. Coding artifacts include features related to the detection of double compression. If an attacker wants to modify a genuine video stored in compressed form, he should decode, edit and then recompress it. Frame deletion in HEVC (High Efficiency Video Coding) coded videos can be detected from the picture type changes [2]. Deletion of the frame results in change in type of frames. This change in type leads to irregularity in coding features of Coding Units/Prediction Units/Transform Units, which are considered as the processing units in HEVC type videos. These features are extracted and then by applying machine learning techniques such as Linear Discriminant Analysis and Multilayer perceptron method, we can classify video as forged or genuine. In [3], generalized variation in prediction footprint is used for detection of recompression and estimation of size of group of pictures. However, the method applies only to MPEG-4/MPEG-2/H.264 video coding standards.

Integrity of video file is verified by analyzing the structure and behavior of video containers generated by mobile devices shared through instant messaging applications, social network and editing software [4]. Using atom extraction method specific features are extracted and classification is done using machine learning approaches such as t-Distributed Stochastic Neighbor Embedding, Pearson correlation coefficient and Principal Component Analysis. In [5], tampering detection is done using unsupervised analysis making use of dissimilarity between original and processed containers. The disadvantage is that the method is applicable only to MP4 file format. In [6], video integrity verification based on blockchain framework is introduced. This method uses Elliptic Curve Integrity Encryption Scheme (ECIES) and Hash-based Message Authentication Code (HMAC) for verification of integrity of videos. Video segments are key-hashed and stored in a chain in chronological order. For verification hash value of the video segment is computed and compared with the hash in the blockchain. In [6], if the verifier himself is an attacker, then he can easily decrypt the key value leading to a collision attack.

## 2. Proposed Method

The video integrity verification method proposed uses the benefits of blockchain framework. This method does not consider the type of frames or coding parameters such as motion vectors and quantization parameters. Our method is applicable to any type of video and can detect any type of forgeries. Video clips recorded every few minutes are termed as video segments. Signature of a video segment is stored in the corresponding block and signature of the whole block is stored in next block. The blocks are chained in order in which videos are captured using Closed Circuit Television, Accident Data Recorder, etc. The overall structure of our method is shown in Figure 1. The private keys for both block signature and video segment signature are randomly generated. For generation of signatures Elliptic Curve Digital Signature Algorithm (ECDSA) [7] is used. ECDSA is faster than other digital signature algorithms based on integer factorization and discrete logarithm and the key length is shorter for providing the same level of security [8]. Section 2.1 details key generation, signature generation and integrity verification using ECDSA.



**Figure 1.** A schematic block diagram of proposed mechanism.

### 2.1. Digital Signature Calculation Using ECDSA

Consider an elliptic curve  $C$  having the equation  $y^2 = x^3 + Ax + B$  defined over a finite field  $F_q$  such that  $A, B \in F_q$ .  $G$  generates the curve  $C$  and  $n$  is the prime modulus, such that the prime power  $q = n^r$  where  $r$  is a large positive integer.

- **Key generation:** Let  $a$  be the private key in the interval  $[1, n-1]$ . Public key  $Q$  is computed by

$$Q = aG \quad (1)$$

- **Generation of signature:** Select a random value  $k$  in the interval  $[1, n-1]$ . Point  $P$  can be calculated using

$$P = kG \quad (2)$$

$P$ 's x-coordinate represents  $R$ . Calculate  $S$  by

$$S = k^{-1}(Z + aR) \bmod n \quad (3)$$

$Z$  represents the hash value. Signature created is  $(R, S)$ . If  $R$  or  $S$  is zero, repeat the generation of signature with different random number  $k$ .

- **Verification of signature:** The signature is invalid if  $R$  and  $S$  are values not in the interval  $[1, n-1]$ . The point  $P'$  is computed by

$$P' = S^{-1}ZG + S^{-1}RQ \quad (4)$$

If  $R$  and x-coordinate of point  $P'$  are equal, signature is valid. Otherwise signature is invalid. If  $P' = 0$ , signature is rejected.

## 2.2. Proposed Blockchain Generation

Each block includes signature of video segment, signature of previous block, public key of video segment signature, public key of previous block signature, path of the video file, and index of the block. The following process describes blockchain generation.

1. Compute the public keys  $Q_V$  and  $Q_B$  using Eq. (1) from corresponding private keys  $a_V$  and  $a_B$  respectively, where  $Q_V$  and  $a_V$  are the keys of the video content and  $Q_B$  and  $a_B$  are the keys of previous block.
2. Generate the random values  $k_V$  and  $k_B$  for the video segment and previous block respectively.
3. Generate the hash values  $Z_V$  and  $Z_B$  for video and previous block respectively using SHA-512 algorithm.
4.  $P_V$  and  $P_B$  are the points for the video segment and previous block respectively, which can be calculated by Eq. (2).  $R_V$  and  $R_B$  represent the x-coordinates of  $P_V$  and  $P_B$  respectively.
5. Calculate  $S_V$  and  $S_B$  corresponding to video content and previous block using Eq. (3).  $(R_V, S_V)$  is the signature of video segment,  $(R_B, S_B)$  is previous block's signature.
6. Add block  $B_V$  containing fields  $((R_V, S_V), (R_B, S_B), Q_V, Q_B, Index, Vname)$  into the blockchain.

## 2.3. Video Integrity Check

Integrity verification of the particular block  $B_V$  as follows:

1. Extract the block  $B_V \rightarrow ((R_V, S_V), (R_B, S_B), Q_V, Q_B, Index, Vname)$ .
2. Compute  $Z_B'$ , the hash of previous block of  $B_V$  and  $Z_V'$  is the hash of the video segment  $Vname$  from  $B_V$ .
3. Compute  $P_B'$  and  $P_V'$ , points corresponding to the verification of previous block and video content using Eq. (4).
4. If x-coordinate of  $P_B'$  is equal to  $R_B$ , then previous block signature is verified. Likewise x-coordinate of  $P_V'$  is equal to  $R_V$ , then video content signature is verified. If both the signatures are valid the video is genuine otherwise tampered with.

## 3. Experimental Results

Five video segments [9], [10], [11], [12], [13], which are publicly available are used in the experiments. These five videos are of resolution 1280x720 pixels and frame rate is 30 frames per second. Forged videos are created by using AVS video editor [14]. Tampered videos are created by deletion/copy-paste/insertion of frames in the video

segment. Experiments were conducted on a PC having Intel Core i7-45U CPU@1.8GHz×4 and 12 GB RAM. The test videos were encoded using H.264/AVC video codec by FFmpeg [15]. For using cryptographic functions, cryptographic library called OpenSSL is used. Some of the videos from benchmark datasets such as VIRAT [16], SULFA [17], Derf's collection [18] are also used for testing.

### 3.1. Performance Evaluation

Table 1 shows the comparison of performance of our method with best existing method on different test videos with varying sizes. Time to create single block is defined as encoding time and that of verifying single block is verification time. From Table 1 it is obvious that our method is about 35 percent faster than the compared work.

Table 2 compares our method with other video integrity verification/forgery detection methods, where both our method and [6] possess the same capabilities. However, if the case arises that the verifier himself is an attacker, he can easily decrypt the key and hash from the verification integrity code stored in each block leading to a collision attack. In our method, even if the verifier is the authorized person, he can only verify the signature is valid or not.

**Table 1.** Encoding and verification time of proposed method in comparison to existing blockchain method

Video	Encoding time (Seconds)		Verification time (Seconds)	
	ECIES with HMAC	ECDSA	ECIES with HMAC	ECDSA
Video_1 [9]	0.0122	0.0080	0.0120	0.0077
Video_2 [10]	0.0184	0.0119	0.0184	0.0117
Video_3 [11]	0.0860	0.0567	0.0860	0.0565
Video_4 [12]	0.1650	0.1100	0.1630	0.1090
Video_5 [13]	0.0428	0.0270	0.0420	0.0269
VIRAT_S_010005_02_000177_000203 [16]	0.0138	0.0090	0.0136	0.0088
VIRAT_S_010106_03_000730_000782 [16]	0.0480	0.0310	0.0475	0.0305
VIRAT_S_000200_01_000226_000268 [16]	0.0188	0.0123	0.0185	0.0121
08_original [17]	0.0167	0.0108	0.0161	0.0105
05_original [17]	0.0324	0.0210	0.0316	0.0206
harbor_cif [18]	0.0493	0.0323	0.0488	0.0320
football_422_ntsc [18]	0.1826	0.1201	0.1807	0.1186

**Table 2.** Comparison of integrity verification capabilities of different methods.

Method	Dependency on file type	Inter-frame tampering detection	Intra-frame tampering detection
[2]	Yes (HEVC only)	Yes (Deletion only)	No
[3]	Yes (MPEG-2, MPEG-4, H.264 only)	Yes	No
[4]	Yes (MP4, MOV, 3GP only)	Yes	Yes
[5]	Yes (MP4 only)	Yes	Yes
[6]	No (Applicable to all types)	Yes	Yes
Proposed	No (Applicable to all types)	Yes	Yes

### 3.2. Security Analysis

By using blockchain, modification in the chained hash leads to integrity violation [19]. In our method, modification of the hash value is prevented by using the digital signatures for the video content and entire block. Two unique private keys are used for each block and these keys are randomly generated. ECDSA is based on infeasibility of solving elliptic curve discrete logarithm problem. From the public key in ECDSA it is difficult to calculate the private key [8]. This difficulty is doubled because of the use of two random keys. For a point  $G$  on  $C$  over  $F_q$  and integer  $a$ , the point  $aG$  can be computed in  $O((\log a)(\log q)^3)$  bit operations. Thus time complexity of digital signature calculation is  $O((\log a)(\log q)^3)$ . Given  $G$  and  $aG$ , fastest known algorithms can compute  $a$  in  $O(\sqrt{q})$  [20], which is greater than the time complexity for solving integer factorization problem. Furthermore, a digital signature only verifies the signature is valid or not, and does not reveal the content.

- **Brute-Force Attack:** One has to guess both block signature and video content signature for each block. Computational complexity is exponential in terms of key length [21] to find the key for each signature in each block. Unique pair of keys is used to generate signature for each video segment and each block which makes brute force attack even more difficult [8].
- **Collision Attack:** One tries to determine two messages having same value of hash. To attack the hash function SHA-512, the number of computations is of the order of  $2^{m^2}$ ,  $m$  is output size in bits [21]. If initialization vector and algorithm are known, hash code for the messages can be generated to find the collision [21]. However, for ECDSA, the private key also should be known to create a message and code pair. For a successful attack, he should know both private keys for block and video content signature which is infeasible.

## 4. Conclusion

We proposed a blockchain based method for integrity verification of video data. Each block in blockchain includes the digital signature of the video content and previous block. Blocks are chained in chronological order. In verification, validity of both the signatures is checked. From the experimental result, it is evident that our method applies to any video type and detects any forgery type. Also, the analysis of the computational complexity confirms that the proposed method is faster. Use of ECDSA reduces memory consumption as the key size is smaller and provides higher level of security in comparison to other digital signature algorithms. Security analysis demonstrates that integrity verification method proposed is more resistant to brute force attacks and collision attacks. The proposed method can be used for video integrity verification in applications which demand faster integrity verification and higher robustness against tampering attempts.

## References

- [1] Singh RD, Aggarwal N. Video content authentication techniques: a comprehensive survey. *Multimedia Systems*. 2018; 24: p. 211–240.

- [2] Hong JH, Yang Y, Oh BT. Detection of frame deletion in HEVC-coded video in the compressed domain. *Digital Investigation*. 2019; 30: p. 23–31.
- [3] Vázquez-Padín D, Fontani M, Shullani D, Pérez-González F, Piva A, Barni M. Video integrity verification and GOP size estimation via generalized variation of prediction footprint. *IEEE Transactions on Information Forensics and Security*. 2019; 15: p. 1815–1830.
- [4] Huamán CQ, Orozco ALS, Villalba L.J.G. Authentication and integrity of smartphone videos through multimedia container structure analysis. *Future Generation Computer Systems*. 2020; 108: p. 15–33.
- [5] Iuliani M, Shullani D, Fontani M, Meucci S, Piva A. A video forensic framework for the unsupervised analysis of MP4-like file container. *IEEE Transactions on Information Forensics and Security*. 2018; 14: p. 635–645.
- [6] Ghimire S, Choi JY, Lee B. Using blockchain for improved video integrity verification. *IEEE Transactions on Multimedia*. 2019; 22: p. 108–121.
- [7] Al-Zubaidie M, Zhang Z, Zhang J. Efficient and secure ECDSA algorithm and its applications: a survey. *arXiv preprint arXiv:1902.10313*. 2019.
- [8] Jana B, Poray J. A performance analysis on elliptic curve cryptography in network security. In 2016 International Conference on Computer, Electrical & Communication Engineering (ICCECE); 2016. p. 1–7.
- [9] Real yellow car-YouTube. [Online]. [Accessed on April 2021]. Available from: <https://www.youtube.com/watch?v=o2YNaYcwdbA/>.
- [10] Nature in 30 seconds-YouTube. [Online]. [Accessed on April 2021]. Available from: <https://www.youtube.com/watch?v=MHna8CzxPLk/>.
- [11] 1 min of nature footage—4K (Ultra HD)-YouTube. [Online]. [Accessed on April 2021]. Available from: <https://www.youtube.com/watch?v=WLKJnHu0GC4/>.
- [12] 4K video Ultra HD—Epic footage-YouTube. [Online]. [Accessed on April 2021]. Available from: <https://www.youtube.com/watch?v=od5nla42Jvc/>.
- [13] 029-Realistic beautiful flower painting timelapse by artistbrownlion— Satisfying video—2.5 Min-YouTube. [Online]. [Accessed on April 2021]. Available from: [https://www.youtube.com/watch?v=2g8bS\\_nNYE&t=7s/](https://www.youtube.com/watch?v=2g8bS_nNYE&t=7s/).
- [14] AVS video editor. [Online]. [Accessed on Dec. 5, 2020]. Available from: <https://www.avs4you.com/avs-video-editor.aspx/>.
- [15] FFmpeg. [Online]. [Accessed on Nov. 20, 2020]. Available from: <http://www.ffmpeg.org/>.
- [16] Oh S, Hoogs A, Perera A, Cuntoor N, Chen CC, Lee JT, et al. A large-scale benchmark dataset for event recognition in surveillance video. In CVPR 2011; 2011. p. 3153–3160.
- [17] Qadir G, Yahaya S, Ho ATS. Surrey university library for forensic analysis (SULFA) of video content. 2012.
- [18] Derf's Collection. [Online]. [Accessed on August. 6, 2021]. Available from: <https://media.xiph.org/video/derf/>.
- [19] Aitzhan NZ, Svetinovic D. Security and privacy in decentralized energy trading through multi-signatures, blockchain and anonymous messaging streams. *IEEE Transactions on Dependable and Secure Computing*. 2016; 15: p. 840–852.
- [20] Calabresi M. An introduction to elliptic curve cryptography. Ohio State Univ. 2016;: p. 10.
- [21] Stallings W. Cryptography and network security, 4/E: Pearson Education India; 2006.

# High Payload Qr-Based Data Hiding Using Secured Compressed Watermark in Polar Domain

Indrarini Dyah IRAWATI<sup>a,1</sup>, Gelar BUDIMAN<sup>b</sup>, Kholidiyah MASYKUROH<sup>c</sup>, Zein Hanni PRADANA<sup>c</sup> and Arfianto FAHMI<sup>b</sup>

<sup>a</sup>*School of Applied Science, Telkom University, Indonesia*

<sup>b</sup>*School of Electrical Engineering, Telkom University, Indonesia*

<sup>c</sup>*Fakultas Teknik Telekomunikasi dan Elektro, Institut Teknologi Telkom Purwokerto, Indonesia*

**Abstract.** Audio Watermarking is a method to insert a copyright marker on audio. This method inserts a watermark in the information form and in a way that does not damage the audio. This technique is one of the ways to solve the problem of copyright infringement. The embedded watermark has to meet the condition of not damaging the audio and must have robustness, imperceptibility, and good capacity. The data hiding technique use the combined method of Lifting Wavelet Transform (LWT), Fast Fourier Transform (FFT), QR Decomposition and Reconstruction, and Cartesian-Polar Transformation (CPT) based on Quantization Index Modulation (QIM) with the secured and compressed watermark using Compressive Sampling (CS) technique. The proposed scheme is blind Audio Watermarking as it no needs for original audio in the detection process. The combination of methods overcomes multiple attacks with guaranteed quality watermarking and high capacity. Compared to the existing technique, the data hiding technique can withstand LPF attacks, Resampling, Linear speed change (LSC), and MP3 compression. This proposed technique is also secured due to the coded watermark by a particular random key using CS. Combining CS and Audio Watermarking techniques can perform well in capacity, imperceptibility, security, and attack resistance.

**Keywords.** Audio Watermarking, CPT, CS, FFT, LWT, QR Decomposition, QIM

## 1. Introduction

Digital watermarking is inserting digital watermarks on information such as audio signals, images, or documents used as copyright markers [1]. The insertion of a watermark on the information must not destroy the original data. Audio Watermarking is an application that is compatible with the Human Auditory System (HAS). In research [2], Audio Watermarking schemes can maintain copyright security and are robust against

---

<sup>1</sup> Corresponding author, School of Applied Science, Telkom University, Indonesia; E-mail: indrarini@telkomuniversity.ac.id.

various standard signal processing attacks such as noise addition, filtering, quantization, echo, invert, and compression.

Audio Watermarking research has been developed by several previous researchers [2-17]. Singha and Ullah propose a combination of multi-level Discrete Wavelet Transform (DWT) and Singular Value Decomposition (SVD) on multiple images of different sizes as a decentralized watermark [3]. Chen and Wornell used the Quantization Index Modulation (QIM) technique which has good performance on the parameters of embedding robustness. The degradation of the host signal can solve by removing interference [4].

Research [5-9] exploited an Audio Watermark using the Lifting Wavelet Transform (LWT) method. In [5], Hu and Lee explored LWT to level 3 decomposition for blind audiowatermarks capable of overcoming cropping and alternation attacks, while effectively detecting tamper and self-recovery. Another study combines audio watermarking techniques, namely Cordic QR decomposition and multi-resolution decomposition to produce a good balance of perceptual transparency, robustness, and payload. [6]. A study by Lei combines the LWT-DCT-SVD method. The method has resistant to hybrid and desynchronization attacks with good imperceptibility [7]. Budiman [8] proposed Audio Watermarking using QR and CPT in the SWT-DST domain with high imperceptibility and good robustness to MP3 compression and resampling. However, its payload is still low or less than 100 bps. LWT with the Spread Spectrum method and has the best optimization results can only be done on optimized audio, has poor robustness against some attacks. In the study [9], the authors combined the LWT-QR factorization based on QIM. The correlation coefficient result can increase the security of the watermarking algorithm.

The DWT technique is also applied to Audio Watermarking by the authors [3, 10-11]. For example, paper [10] proposes a hybrid watermarking algorithm for voice signal cryptography to protect the copyright of audio files. The algorithm is formed by combining three-domain transformations, DCT, DWT, and SVD, increasing the strength of resistance to audio processing attacks. Another study combined second-level 2D-Haar DWT and Schur decomposition. This hybrid scheme has succeeded in increasing imperceptibility, robustness, and data payload size [11].

In [12], the authors proposed a combined transform LWT-DCT, then decomposed by QR, then embedded by QIM. The result is good imperceptibility and high capacity, but the robustness is still moderately low against LPF and resampling attacks. Looking for performance differences between FFT and DCT using the Fibonacci sequence of numbers to generate high capacity, deliver robustness against common signal processing attacks, and cause very low perceptual distortion. Irawati et al. combined audio watermarking techniques based on QIM with LWT, DCT, QR Decomposition. The combined performance of the methods results in a powerful and invisible audio watermark with a very high watermark payload [13]. In [14], the authors presented an audio watermarking with a CS scheme. The audio host is embedded by the watermark and compressed by the CS technique using a multi-bit spread spectrum (SS) framework. Research by Kaur et al. inserted watermarking bits using the embedding function on each frame in the  $R$  matrix [15]. Patel and Amin performed CS for optimizing signal reconstruction in Audio Watermarking [16]. CS samples the image before embedding the image into the audio host by QIM. The audio host is represented as a one-dimensional matrix. Then LWT decomposes a one-dimensional matrix into some sub-bands. The low-frequency sub-band will be further processed. The DCT technique converts the signal

into essential frequency components. The QR factorization decomposes each DCT frame into an orthogonal matrix ( $Q$ ) and an upper triangle ( $R$ ).

In this study, we proposed a secured-compressed watermark system based on QR. The proposed method is a combination of several techniques. First, a host is decomposed with LWT, and then the signal is converted into the frequency domain with FFT. Next, the signal converted into the frequency domain is converted into the  $M \times M$  matrix to retrieve the  $R$  matrix with a QR decomposition. After the signal becomes a matrix, each element in the matrix will be changed into a polar form using the Cartesian-Polar Transform (CPT). The watermark image is altered from a two-dimensional state to a one-dimensional structure, and after that, the CS process 1 is the same as in [17] to obtain watermark compression. After the watermark is pre-processing and CS, each watermark bit is inserted into each matrix element using QIM. After all the watermark bits are inserted, the inverse process in ICPT is carried out to return the bits in the  $R$  matrix element into cartesian form and proceed by reconstructing the  $R$  and  $Q$  matrix with QR reconstruction. After that, the  $M \times M$  matrix is converted back into a 1-dimensional signal again by IFFT into a back-time domain. Finally, ILWT changes the reconstructing of the signal into complete audio again. LWT decomposes the audio in the extraction process, then converted into the frequency domain by FFT. The 1-dimensional signal is converted into an  $M \times M$  matrix to get the  $R$  matrix from QR decomposition. Then CPT is performed to convert each element into a polar shape. Then after the CPT is carried out on each component in the  $R$  matrix, the extraction process is carried out by QIM. QIM extracted CS reconstructed bits, and post-processing was carried out to obtain the perfect watermark shape. Those methods are expected to result in high robustness to noise addition, cropping, resampling, signal addition, signal reduction, and MP3 compression.

The next section of the paper is described as follows: section 2 illustrates the basic theory of Audio Watermarking, transformation technique, and CS. Section 3 explained the embedding and extraction process of the Audio Watermarking system. While section 4 presented the analysis result of the proposed Audio Watermarking system with scenarios before and after attacks. The last section is the conclusion and future works.

## 2. Basic Theory

This paper used LWT to select bits as the selection point. FFT method to change the domain host to the frequency domain, QR Decomposition method to build the matrix, Cartesian Polar Transform method to change the polar form to cartesian form. CS method to compress the watermark before embedded, and Quantization Index Modulation to embed the watermark.

### 2.1. Lifting Wavelet Transform

LWT is faster than traditional Wavelets and can generate unsuitable resolution analyses on a uniform grid [18,19]. The Lifting Wavelet scheme process is divided into three steps: split, predict, and update.

- *Split*

Split is a step to divide the audio host signal that has been segmented into two-part, where the initial sequence is the odd part, and the final half order is the even part, as the following equation explains both sections.

$$x[n] = x_e[n] + x_o[n] \quad (1)$$

$$x_e[n] = x[2n] \quad (2)$$

$$x_o[n] = x[2n + 1] \quad (3)$$

where  $x[n]$  is the segmented audio signal, notation  $x_e[n]$  does the even part take the segmented signal,  $x_o[n]$  is the segmented signal taken by the odd part, and  $n$  is the time domain.

- *Prediction*

After the signal is divided into two, then the prediction is made by using even sequence correlation and odd sequence shown by Wavelet Coefficient ( $d[n]$ ). Wavelet Coefficient is obtained by calculating the original value (odd) minus the prediction value (taken from the even part). The following equation explains the Prediction process.

$$d[n] = x_o[n] - P(x_e[n]) \quad (4)$$

where  $d[n]$  is the coefficient of Wavelet and  $P(x_e[n])$  is the prediction operator of  $x_e[n]$ .

- *Update*

All subsets of split results have values that may differ from the original data. However, the general characteristics of the data must be kept consistent through the update process. The following equation describes the update process.

$$c[n] = x_e[n] + U(d[n]) \quad (5)$$

where  $c[n]$  is the lifted loop, and  $U(d[n])$  is the prediction operator.

## 2.2. Fast Fourier Transform

The FFT algorithm is based on the principal principle of the decomposition of the discrete Fourier transform (DFT) calculation of a sequence along  $N$  into the DFT.

The FFT is defined by the formula [20]:

$$H(k) = \sum_{n=0}^{\left(\frac{N}{2}\right)-1} h(2n)W_{N/2}^{nk} + \sum_{n=0}^{\left(\frac{N}{2}\right)-1} h(2n+1) \quad (6)$$

where  $H(k)$  is the domain transformation value, the digital media value is notated by  $h(n)$ ,  $k$  is the frequency domain,  $n$  is the time domain, and  $N$  is the amount of data converted into the frequency domain.

To perform inversion Fast Fourier Transform (IFFT) then it can do with Eq. (7):

$$h(n) = \frac{1}{N} \sum_{n=0}^{\left(\frac{N}{2}\right)-1} \left[ Re * \cos\left(\frac{2\pi kn}{N}\right) + Im * \sin\left(\frac{2\pi kn}{N}\right) \right] \quad (7)$$

where  $Re$  being an actual number,  $Im$  is an imaginary number, and the serial value of the watermark complex is notated by  $h(n)$ .

### 2.3. QR Decomposition

The QR Decomposition of the quadratic matrix  $A$  is decomposition  $A$ , which is defined by Eq. (8) [21,22]:

$$A = Q \cdot R \quad (8)$$

where  $Q$  is an orthogonal matrix with  $Q^T \cdot Q = I$  and  $R$  are an upper triangular matrix.

If  $A$  is a nonsingular matrix, then factorization is unique. There are several methods for calculating QR decomposition. One such method is the Gram-Schmidt process. Based on the Gram-Schmidt procedure, the matrix column  $A$  is considered as the processed vector as shown in (9).

$$A = [a_1 | a_2 | \dots | a_n] \quad (9)$$

Based on the Gram-Schmidt process then the result of QR factorization is as follows.

$$A = [a_1 | a_2 | \dots | a_n] = [e_1 | e_2 | \dots | e_n] \begin{bmatrix} a_1 \cdot e_1 & a_2 \cdot e_1 & \dots & a_n \cdot e_1 \\ 0 & a_2 \cdot e_2 & \dots & a_n \cdot e_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_n \cdot e_n \end{bmatrix} = Q \cdot R \quad (10)$$

where  $a_1 \dots a_n$  is a column of matrices  $A$  and an orthonormal set in column matrices indicated in the notation  $e_1 \dots e_n$ .

### 2.4. Cartesian-Polar Transform

In the polar coordinate system  $(r, \theta)$ ,  $r$  shows the distance from the origin, and  $\theta$  shows the angle between the reference line and the line through the origin and endpoints. The transformation to polar coordinate is given by the Eq. (11) [8]:

$$r = \sqrt{x^2 + y^2}, \quad \theta = \tan^{-1} \left( \frac{y}{x} \right) \quad (11)$$

where  $(x, y)$  is a point on the polar coordinate system. Transformations from polar coordinates to cartesian coordinates are given in Eq. (12)

$$x = r \cos \theta, \quad y = r \sin \theta \quad (12)$$

### 2.5. Cartesian-Polar Transform

Quantization Index Modulation (QIM) is a technique of inserting a Watermark depending on a signal by quantizing a host signal with a selected quantization based on the bit value of the message encoded.

In QIM, there is parameter  $\Delta$  used as an insertion parameter. This parameter is based on the quantity of  $n$  bit or bits quantization in the QIM method by calculating the distance between quantization. The following are the equations used for the insertion process and the extraction process on the watermark.

- QIM embedding process:

If  $w = 1$  then,

$$F'(0) = A_k \text{ and } \arg \min |F(0) - A_k| \quad (13)$$

If  $w = 0$  then,

$$F'(0) = A_k \text{ and } \arg \min |F(0) - A_k| \quad (14)$$

where,

$$A_k = \left(2k + \frac{1}{2}\right)\Delta, B_k = \left(2k - \frac{1}{2}\right)\Delta, k = 0, \pm 1, \pm 2, \dots \dots \quad (15)$$

- QIM extraction process:

$$\tilde{w}(k) = \text{mod} \left( \text{ceil} \left( \frac{F(0)}{\Delta} \right), 2 \right) \quad (16)$$

$$\Delta = \frac{1}{2^{(nbit-1)}} \quad (17)$$

where  $F(0)$  is the audio host sample before it is quantized,  $F'(0)$  is the audio host sample after embedded.  $w$  is watermark bit before embedded, watermark bit after embedded or before extraction shown in the variable  $\tilde{w}$ . Variable  $\Delta$  is delta or the distance between the quantization values. Variable  $k$  is the quantization index, and  $n$  bit is bit quantization on the QIM method.

## 2.6. Compressive Sensing

CS is a new signal processing paradigm. CS is also a technique that can reconstruct data from incomplete measurements. CS utilizes a unique structure in the data, namely data sparsity. The sparse data is represented with a few non-zero values [17].

A signal expressed in  $R^N$  is represented in a vector base  $N \times 1 \{\psi_i\}_{i=1}^N$  An orthogonal form. Using the  $N \times N$  base matrix  $\psi$  with  $\{\psi_i\}_{i=1}^N$  like the column, the actual value of the dimension signal  $f$  can write with the following equation.

$$f = \psi x \quad (18)$$

An  $x$  is said to be “K-sparse” only if  $K$  comes from the coefficient  $x_i$ , which is a non-zero coefficient when  $(N - K)$  is zero, then  $K \ll N$ . If the linear measurement process is general, then  $M \times N$  on the matrix  $\Phi$  ( $M \ll N$ ) will satisfy.

$$y = \Phi f \quad (19)$$

where  $y$  is the vector  $M \times 1$ , the measurement vector, and  $\Phi$  is the matrix  $M \times N$ . Based on the first Eq 1, it can be formulated into

$$y = \Phi f = \Phi \psi x = \Theta x \quad (20)$$

where  $\Theta = \Phi \psi$  is an  $M \times N$  matrix known as the CS matrix. As for CS reconstruction with the minimum of  $\ell_1$  can be shown in the following formulas [17].

$$\min_x \|x\|_{l1} \text{ subject to } y = \Phi x \quad (21)$$

where  $l1$ -norm of the signal  $x$  is the absolute value of the sum of the signal samples and is expressed in the following equation:

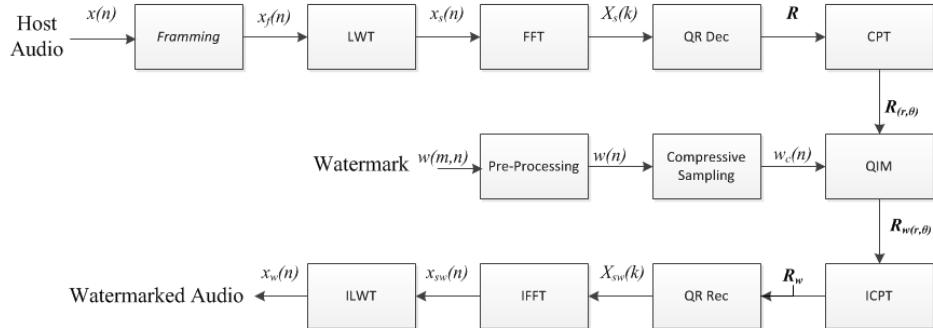
$$\|x\|_{l1} = \sum_{i=1}^N |x_i| \quad (22)$$

### 3. Watermarking Model

This section explains the modeling system used, which begins with the embedding process. Various attacks will attack it after it does the extraction process, and then it calculates the quality output of the Audio Watermarking.

#### 3.1. Embedding Process

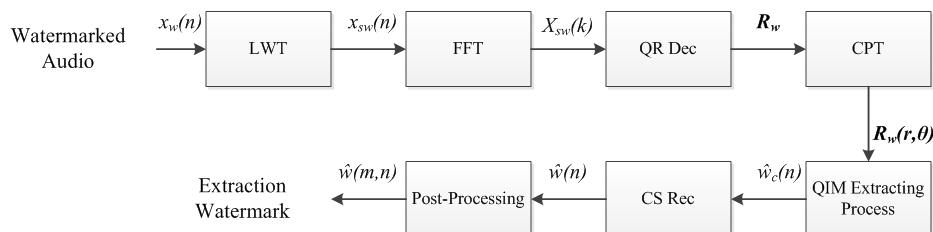
The embedding process has three steps. There is a bit selection process as insertion place, host domain transformation, and bit watermark insertion process. Figure 1 shows the embedding process flow chart.



**Figure 1.** Embedding Process Flow Chart.

#### 3.2. Extraction Process

After we attack watermarked audio, the watermarked audio data will extract. The step of extracting process is indicated in Figure 2:



**Figure 2.** Extraction Process Flow Chart.

#### 4. Simulation Result

In the simulation, we have input parameters such as  $nbit$  as bit number of QIM,  $Nframe$  as sample number per segment, and  $N$  as decomposition level. These parameters affect the output or performance parameters such as ODG and SNR as watermarked audio quality,  $C$  as watermark payload, and BER as watermark robustness. We apply the simulation with the change of parameters to understand the relation between the input parameters and the performance parameters. We use an image for the watermark, with a resolution of 10x15 pixels. We also use four audio files such as piano.wav, voice.wav, guitar.wav, and bass.wav as host files. The host file sampling frequency is 44.1 kHz with 16-bit audio quantization. The simulation results are shown in the graph in the Figure 3, Figure 4, and Figure 5.

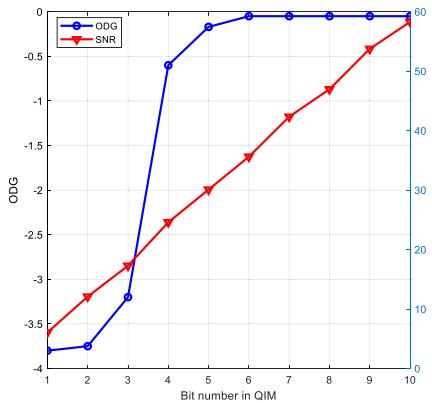


Figure 3. The  $n$  bit effect ODG and SNR.

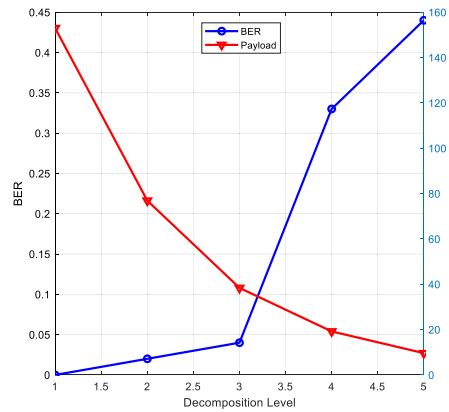


Figure 4. The  $N$  effect to BER and payload.

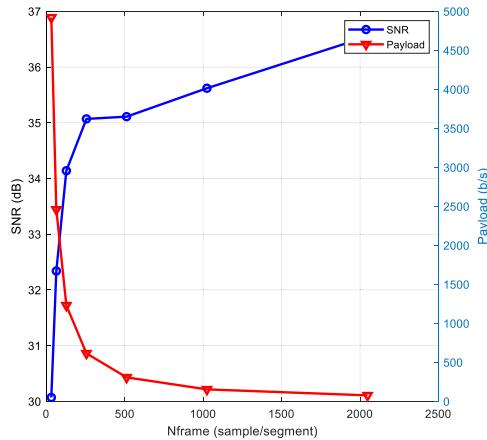


Figure 5. The  $Nframe$  effect to SNR and payload.

The first simulation with voice.wav as a host file changes  $nbit$  from 1 to 10 bits while  $N = 1$  and  $Nframe = 1024$  sample/segment. These parameters obtain the watermark payload  $C = 153$  bps with the minimum, and maximum BER is 0 and 0.18, respectively. The audio quality represented by SNR and ODG is displayed in Figure 3. The rise of

*nbit* increases the audio quality in ODG and SNR. This improvement is because of the reduction of modified audio amplitude by QIM when *nbit* increases. However, the PEAQ procedure calculates ODG with the perceptual process. Thus, the rise of ODG is not linear compared to the increase of SNR. For example, analyzing Figure 1, we can select minimum *nbit* = 4 due to the SNR>20 dB and ODG>-1.

The second simulation with voice.wav as a host file then changes decomposition level or *N* from 1 to 5 while *Nframe* = 1024 sample/segment and *nbit* = 6. The simulation result with these parameters shows in Figure 5 and obtains SNR 35.62 to 42.82 dB and ODG -0.05 to -0.80. Figure 4 shows that the trend in *N* is aligned with the BER value and opposite to the watermark payload. The larger *N*, the BER increases but the payload of the watermark decreases. This fluctuation illustrates the trade-off between the durability and payload of the watermark when the decomposition rate changes. We can select the maximum decomposition level *N* = 3 due to BER<10%.

The third simulation with voice.wav as a host file changes *Nframe* from 32 to 2048 while *N* = 1 and *nbit* = 6. BER=0 and ODG 0.08 to -2.59 are the parameters that can obtain. Figure 5 displays the effect of *Nframe* on the watermarked audio quality and the watermark payload. By increasing *Nframe*, the payload decreases but on the other hand, the audio quality or SNR increases. This fluctuation displays the trade-off between the watermarked audio quality and the watermark payload when *Nframe* changes. From Figure 5, we can choose all values of *Nframe* 32 to 2048 since the payload is still more than 20 bps and the audio quality is still more than 20 dB. Thus, we can select a certain *Nframe* that is robust to a specific attack.

**Table 1.** Watermark robustness test with BER parameter.

Attacks	Parameter	BER			
		Voice	Piano	Guitar	Bass
LPF	3k	0	0.011	0.019	0
	6k	0	0.011	0	0
	9k	0	0.011	0	0
BPF	100-6k	0.312	0.195	0.423	0.45
	50-6k	0.093	0.05	0.335	0.253
	25-6k	0.023	0.046	0.328	0.089
Noise	0 dB	0.468	0.511	0.464	0.484
	10 dB	0.464	0.460	0.472	0.437
	20 dB	0.222	0.25	0.238	0.394
Resampling	22.05k	0	0.011	0	0
	11.025k	0	0.011	0	0
	16k	0	0.011	0	0
LSC	1%	0	0.023	0.019	0
	5%	0.031	0	0.011	0
	10%	0.031	0	0.011	0
Equalizer		0.148	0.210	0.312	0.328
MP3 Compression	32k	0.093	0.109	0.148	0.128
	64k	0.058	0.046	0.074	0.027
	128k	0	0	0.019	0.027
	192k	0	0	0.019	0
Average			0.204		

We apply the fourth simulation by parameters  $N = 1$ ,  $N_{frame} = 512$  sample/segment, and  $nbit = 4$ . This parameter yields the average ODG -1.79 and SNR 23.27 dB with payload 307.62 bps. Finally, we apply our proposed method of several attacks, as displayed in Table 1. The watermark robustness testing shows that the method we developed is resistant to various attacks, including LPF, resampling, LSC, and MP3 compression using 128 kbps bit rate.

BER is a watermark resistance test parameter. This value describes the quality of the extracted watermark image from the attacked Audio Watermark. The higher the BER, the worse the quality of the extracted watermark image, and vice versa. There is a maximum BER threshold on the damaged watermark image, but it can still be accepted and understood by humans visually. The resolution of the watermark image that is Table 2 shows the watermark image from the original image in Figure 3. with a resolution of 10 x 15 and BER = 0. The extracted watermark can be read even though the watermark image is damaged in the range of BER = 0.011 to 0.074. While the BER of the watermark image is 12.8% and above, the extracted watermark is unreadable. This condition shows the watermark robustness is acceptable when the BER is up to 7.4%.

**Table 2.** Visualization of Watermark Extraction.

BER	Watermark Extraction	BER	Watermark Extraction
0		0.058	
0.011		0.074	
0.027		0.128	
0.046		0.222	

Table 3 describes performance comparison between our proposed method and the previous method using the QR decomposition as one of the processes before embedding. Our proposed method has relatively better in terms of robustness with high capacity, despite lower capacity than [13]. Nevertheless, the robustness in [13] is too low compared to our proposed method's robustness. NA from the table means not available or not reported. The imperceptibility of our method is the lowest, but it is still accepted due to more than the standard minimum of audio quality or 20 dB [23].

**Table 3.** Performance Comparison.

Ref.	Robustness/BER (%)				Imperceptibility/SNR (dB)	Capacity (bps)
	MP3 64k bps	MP3 128k bps	LPF	Resampling 22.05 kHz		
[8]	9	3	NA	0-1	32.72	21.43
[13]	0-14	0-5	0-14	0-14	31.40	615.23
Proposed	2.7-7.4	0-2.7	0-1.1	0-1.1	23.25	307.62

## 5. Conclusion

We propose the Audio Watermarking technique with the combined Lifting Wavelet Transform, Fast Fourier Transform, QR Decomposition and Reconstruction, and Cartesian-Polar Transformation based on Quantization Index Modulation. The watermark is sampled using CS to reduce the embedded watermark size. The selected parameters from the simulation yield the average value of ODG is -1.79, the average value of SNR is 23.35, and the watermark payload 307.62 bps. Using the selected parameters, the method we propose has a good performance as robust to various attacks, such as LPF, resampling, LSC, and MP3 compression with 128 kbps.

There are still opportunities to re-examine the attacks that have not been successfully anticipated using the proposed method. The more technology develops, the more attacks will appear, so it is necessary to explore the features to resist them. In addition, it is required to develop research on the effect of using CS on Audio Watermarks.

## Acknowledgment

This research is supported by Telkom University, Telkom Institute of Technology Purwokerto, and Telkom Foundation.

## References

- [1] Abdulla NTB and Navas KA. High security watermarking techniques for digital rights management: a review. In: International Conference on Communication and Signal Processing (ICCSP); 2020 July 28-30; Chennai, India: IEEE; c2020. p. 162-166.
- [2] Nair U and Birajdar GK. Compressed domain secure, robust and high-capacity audio watermarking. *Iran Journal of Computer Science*. 2020; 3:217–232.
- [3] Singha A and Ullah MA. Development of an audio watermarking with decentralization of the watermarks. *Journal of King Saud University –Computer and Information Sciences*. 2020 September; 1-7.
- [4] Chen B and Wornell GW. Preprocessed and postprocessed quantization index modulation methods for digital watermarking. In: Proceeding Security and Watermarking of Multimedia Contents II; 2000 May 9; San Jose, CA, United States: 3971.
- [5] Hu H and Lee T. Hybrid blind audio watermarking for proprietary protection, tamper proofing, and self-recovery. *IEEE Access*. 2019 December; 7:180395-180408.
- [6] Gupta A, Kaur A, Dutta MK and Schimmel J. Perceptually transparent & robust audio watermarking algorithm using multi resolution decomposition & cordic qr decomposition. In: International Conference on Telecommunications and Signal Processing (TSP); 2019 July 1-3; Budapest, Hungary: IEEE; c2019. p. 313-317.
- [7] Lei B, Soon IY and Tan E. Robust SVD-based audio watermarking scheme with differential evolution optimization. *IEEE Transactions on Audio, Speech, and Language Processing*. 2013; 21(11):2368–2378.
- [8] Budiman G, Suksmono AB, Danudirdjo D and Pawellang S. QIM-based audio watermarking with combined techniques of SWT-DST-QR-CPT using SS-based synchronization. In: International Conference on Information and Communication Technology (ICoICT); 2018 May 3-5; Bandung, Indonesia: IEEE; c2018. p. 286-292.
- [9] Li J and Wu T. Robust audio watermarking scheme via QIM of correlation coefficients using LWT and QR decomposition. In: International Conference on Informative and Cybernetics for Computational Social Systems (ICCSS); 2015 August 13-15; Chengdu, China: IEEE; c2015. p. 1–6.
- [10] Rezaei A, Khalili M. A robust blind audio watermarking scheme based on DCT-DWT-SVD. In Montaser Kouhsari, S, editor. *Fundamental research in electrical engineering*. Singapore: Springer. 2019. p. 101–113.
- [11] Karajeh H, Khatib T, Rajab L and Maqableh M. A robust digital audio watermarking scheme based on DWT and Schur decomposition. *Multimedia Tools Application*. 2019; 78:18395–18418.

- [12] Budiman G, Suksmono AB and Danudirdjo D. CPT-based data hiding in selected subband using combined transform and decomposition method. In: International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC). 2018 December 5-7; Bandung, Indonesia: IEEE; c2019. p. 86-92.
- [13] Irawati ID, Budiman G and Ramdhani F. QR-based watermarking in audio subband using DCT. In: International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC); 2018 December 5-7; Bandung, Indonesia: IEEE; c2019.p. 136-141.
- [14] Budiman G, Suksmono AB and Danudirdjo D. Compressive sampling with multiple bit spread spectrum-based data hiding. Applied Sciences. 2020 June;10(12): 4338-4359.
- [15] Kaur A, Dutta MK, Soni KM and Taneja N. A high payload audio watermarking algorithm robust against Mp3 compression. In: 7th International Conference on Contemporary Computing (IC3); 2014 August 7-9; Noida, India: IEEE; c2014. p. 531–535.
- [16] Patel R and Amin JD. Alpha channel base data hiding using compressive sensing. International Journal of Advance Research and Innovative Ideas in Education. 2016; 3:544–549.
- [17] Huang H, Chang F, Lu Y and Pang Y. Multiple watermarking for compressed sensing with robust transmission applications. In: Global Conference on Consumer Electronics (GCCE). 2018 October 9-12; Nara, Japan: IEEE; c2018. p. 371-372.
- [18] Tomar V, Kumar A and Choudhary A. Conception & implementation of a novel digital image watermarking algorithm using cascading of DCT and LWT. In: International Conference on Reliability, Optimization and Information Technology (ICROIT). 2014 February 6-8; Faridabad, India: IEEE; c2014. p. 501–505.
- [19] Ke WY and Jian W. Hybrid domain audio watermarking embedding algorithm based on redundant discrete wavelet transform. In: International Conference on Computer Systems, Electronics and Control (ICCSEC); 2017 December 25-27; Dalian, China: IEEE; c2018. p. 1134-1137.
- [20] Neyman SN, Pradnyana INP and Sitohang B. A new copyright protection for vector map using FFT-based watermarking. TELKOMNIKA (Telecommunication Comput. Electron. Control). 2014; 12(2):367–378.
- [21] Mohsenfar SM, Mosleh M and Barati A. Audio watermarking method using QR decomposition and genetic algorithm. Multimed. Tools Appl. 2015; 74(3): 759–779.
- [22] Kavitha KJ and Shan BP. Implementation of DWM for medical images using IWT and QR code as a watermark. In: Conference on Emerging Devices and Smart Systems (ICEDSS); 2017 March 3-4; Mallasamudram, India: IEEE; c2017. p. 252-255.
- [23] Katzenbeisser S and Petitolas F. Information hiding techniques for steganography and digital watermarking. Artech House Books; 1999; 28(6).

# Object and Traffic Light Recognition Model Development Using Multi-GPU Architecture for Autonomous Bus

Jheanel ESTRADA <sup>a, b</sup> <sup>1</sup>, Gil OPINA JR <sup>b</sup> and Anshuman TRIPATHI <sup>b</sup>

<sup>a</sup> *Technological Institute of the Philippines-Manila, Philippines*

<sup>b</sup> *Energy Research Institute @ Nanyang Technological University, Singapore*

**Abstract.** The autonomous vehicle is both an exciting yet complex field to dig in these past few years. Many have ventured out to develop Level 4 Autonomous Vehicle but up to this point, many issues were still arising about its safety, perception and sensing capabilities, tracking, and localization. This paper aims to address the struggles of developing an acceptable model for object detection in real-time. Object detection is one of the challenging areas of autonomous vehicles due to the limitations of the camera, lidar, radar, and other sensors, especially during night-time. There were various datasets and models available, but the number of samples, the labels, the occlusions, and other factors may affect the performance of the dataset. To address the mentioned problem, this study has undergone a rigorous process of scene selection and imitation to deal with the imbalance dataset, applied the state-of-the-art YOLO architecture for the model development. After the development process, the model was deployed in a multi-GPU architecture that lessens the computational load on a single GPU structure and was tested on a 12-meter fully electric autonomous bus. This study will lead to the development of a usable and safe autonomous bus that will lead the future of public transportation.

**Keywords.** autonomous driving, autonomous vehicle, intelligent transport system, object detection, real-time, daytime, night time.

## 1. Introduction

Intelligent Transport System (ITS) is a combination of technologies, techniques, and processes that can improve public transit or the whole transportation system management [1]. The public sector has greatly been addressed as one of the predominant drivers of ITS. Early improvements show changes in the traffic management center in the urban areas, automated toll collection, traffic signal controls, and satellite-based dispatching systems. Recent improvements in ITS include autonomous vehicle which is a combination of multi-disciplinary fields such as robotics, embedded systems and circuits, sensors, machine learning, artificial intelligence, and others.

Due to the recent advances in these fields, a combination of highly calibrated cameras, lidars, and radars made this possible to monitor the vehicle and the things around its environment. Autonomous Vehicles (AV) are gradually becoming capable to mimic some human driver actions such as maneuver and defining routes. Recently, many

---

<sup>1</sup> Corresponding Author, Jheanel Estrada, Computer Science Department, Technological Institute of the Philippines, Manila, Philippines; E-mail: jheanelestrada29@gmail.com.

car companies announced that autonomous vehicles will be available to society in the upcoming years [2]. Due to its promising benefits, this attracts many stakeholders to invest in the automotive industry. This predicts that in the year between 2020 and 2025, self-driving cars will be seen on the road [3]. However, due to the complexity and factors such as safety, legal issues, and social acceptance, technical issues, and its capability were questioned [4]. In a natural human driver setting in the road, according to the World Health Organization, nearly 1.25 million people die each year relative to traffic crashes and road accidents [5]. Now, autonomous vehicles' priority is to ensure the safety of both the driver and other road elements. An autonomous vehicle is expected to do the same level of safety as a human driver or even exceeding it. To achieve this, a higher level of sensing capabilities, perception, decision, motion, and mission planning is required.

As part of this global mission to improve the trust and safety of autonomous vehicles, this study aims to develop an acceptable object detection and recognition model by developing an acceptable dataset that covers day time, night time and rain conditions; to deploy the model in a real-time manner that prevents huge frame loss and lastly to utilize a multi-GPU architecture.

## **2. Related Literature and Studies**

To stay in-sync with ongoing research and innovation regarding object detection and recognition, a literature survey was deployed investigating the present-day movements in this field. Image processing, machine learning and map-based techniques are some of the identified methods for this task. The most classic way of this task is through image processing, however, though image processing approach is quite straight and uncomplicated compared to other tasks, it requires critical phases such as thresholding and filtering. One of the undesirable conditions in object detection is the fact that even slight miscalculations or deviations from these standard techniques may result to ambiguous outcomes which is highly sensitive to this task. To somehow addressed this problem, machine learning based methods in combination with some algorithms and ample processing techniques will be helpful to prune the misleading directions.

Object Detection is one of the crucial capabilities of an autonomous vehicle together with the adaption of Deep Neural Network (DNN). All the sensors available in the use of AV are irreplaceable especially cameras since they gather widely usable data or texture information regarding an object in the surroundings of an AV. In literature, there is a wide range of studies about the use of cameras and DNN in the object detection process.

To date, object detection using Deep Neural Network (DNN) can be divided into two (2) strategies namely one-stage detection which is Single Shot Detection (SSD) [6] and You Only Live Once (YOLO) [7], and two-stage detection which includes Spatial Pyramid Pooling (SPP) [8] and Region Convolutional Neural Network (R-CNN) [9]. One stage detection mainly focuses on higher-speed object detection and works on a real-time basis. However, when it comes to high-precision object detection, this type of detection shows a slight disadvantage since it gives out regression detection on the pixel feature map of the CNN and gives the classification and detection results. However, in two-stage detection, aside from obtaining the feature map of an image, it generates

proposals through an RPN through a Region-of-Interest (RoI) then gives the detection and classification result [10].

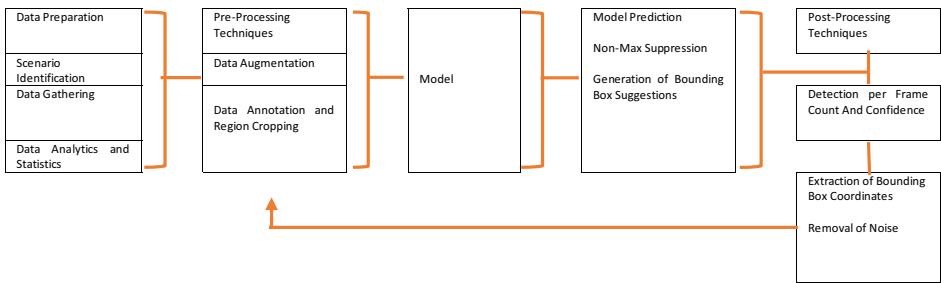
After identifying the type of deep neural network, the next crucial step is to identify the dataset. Today, numerous datasets were available in the market. Each dataset offers advantages over the other in different terms such as the number of classes, labeling techniques, and others. This includes the following:

- Pattern Analysis, Statistical Modelling, and Computational Learning Visual Object Classes (PASCAL VOC) [11];
- Karlsruhe Institute of Technology and Toyota Technological Institute at Chicago Object Detections (KITTI) [12];
- Common Object in Context (COCO) [13];
- and ASTAR3D Dataset that contains 3X the size of nuScenes dataset [13].

One of the crucial steps to obtain optimized results is the collection of large training and testing datasets and the significant amount of time and resources needed to train a good and acceptable model. Moreover, to address the imbalance samples due to inadequate amount of data collected and to provide solutions for image recognition problems, this study will be able to tries to experiment and conduct hardware-based and software-based solutions.

### 3. Methodology

A good object detection model relies on the preparation of dataset and pre-processing methods. Our dataset is comprised of different scenarios taken to imitate the real-world environment. Once the dataset was completed and all the scenarios were generated and imitated, data augmentation processes come in. In the figure 1 below, the over-all process starts with data preparation which includes scenario identification, data gathering, and data analytics and statistics. For scenario identification, risky situations that involve pedestrian and other objects were imitated in a real-world scenario. The examples are the following: pedestrian on the side of the road, pedestrian crossing the road, a pedestrian on the side of the road that does not have the intention to cross, overtaking vehicles, and many more. These scenarios were imitated, staged, and recorded for both daytime and night time. Aside from these scenarios, other common and usual scenarios on the road were recorded. After data gathering, the data that were collected will then be analyzed in terms of statistics and usability. At this point, we are ensuring that all the classes have the same portion of samples. Afterward, the dataset will be pre-processed by region cropping. Region Cropping is a technique we applied to get the usable part of the frame eliminating the top and bottom part of the image. This method is important to lessen the processing time taken for each image since not all the image part is beneficial in the recognition process (see Figure 1 below).



**Figure 1.** Process Diagram from Dataset Preparation to Post-Processing

### 3.1. Dataset

The dataset for this study is comprised of seventeen (17) labels (see Figure 1). The distribution of these labels was also shown in Figure 2. A total of images and approximately 700,000 annotations across labels. The strength of this dataset compared to other famous datasets is it consists of high-density images, heavy occlusions, and a large number of night-time images.



**Figure 2.** Distribution of Samples and Camera Setup

The dataset was developed using 6x FLIR Blackfly BFLY-PGE-31S4M-C mounted on the front side of the bus (see Figure 2 right image). The recording took for a month during both day time and night time along Nanyang Technological University Route. Using Robotic Operating System (ROS), these recorded were stored in a rosbag and extracted ten(10) frames per second. These extracted frames were manually annotated using labelImg tool.

### 3.2. Darknet Framework

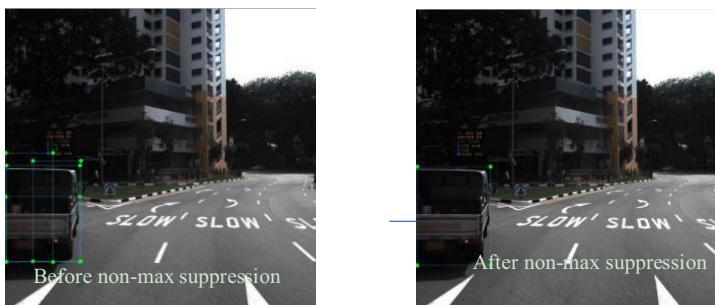
Darknet is an open source framework to train neural networks. It is a framework used to train YOLO networks compatible with Robotics Operating System (ROS) for self-driving cars. YOLO is a powerful neural network in classifying an object using a bounding box around the detected object. It can classify the object in a single pass compared to other implementations which are why it is fast, and performant compared

to other implementations undergo several tasks to classify. Additionally, YOLO works on a real-time basis which is relevant in identifying and recognizing objects on a real-world scenario.

YOLO works by putting and splitting an image into n grid cells (usually 19x19). For each cell that represents a certain part of an object, there will be predicted bounding boxes, confidence scores, and class probabilities. The confidence is calculated using an IOU (intersection over union) metric that measures the overlapping of detected objects with the ground truth as a fraction of the total area of detection. There were many publicly available versions of YOLO. The YOLOv1 was released May 2016, YOLOv2 was released December of the same year, YOLOv3 was released April 2018, YOLOv4 was released May 2020. All versions offer strength on the time they were released. As shown in Figure 3 below, YOLOv3 uses a variant of Darknet which has 53 layers. Because of the task of detection, 53 layers were added on top of the original number of layers that sums up to 106 layers fully convolutional. Aside from these advantages of YOLOv3, we conducted an experiment on our dataset which uses both YOLOv3 and YOLOv4 and it shows that YOLOv3 has a higher mAP (mean average precision) than YOLOv4.

YOLO aims to predict a class of an object in an image by the use of a bounding box (see Figure 3 below). Each bounding box has four (4) descriptors namely:

- Center of bounding box (bxby) or (x, y)
- Width (bw)
- Height (bh)
- C = class or label



**Figure 3. YOLO Bounding Box**

YOLO is predicting that there is an object in the image instead of searching for regions of interest. It is splitting an image  $S \times S$  grid cells. Each cell is responsible for predicting n number of bounding boxes. Each grid cell predicts a bounding box alongside with a confidence value. If a grid cell does not contain a bounding box, its confidence value must be zero. Most of these cells do not contain the object, therefore YOLO will predict a value that will remove boxes with low object probability and bounding boxes with highest. This process is called non-max suppression. If the center of an object falls into a grid cell, that cell is the main responsible for the detection.

Since YOLOv3 is a fully convolutional network, it applies 1x1 kernel on a feature map of three different sizes at three different places in the network. It follows the:

$$1 \times 1 \times (B \times (5+C))$$

Where:

B = number of bounding box a cell on the feature map can predict

C = number of classes

In the case of this study, we used B=3 and C=17, so the kernel size is 1 x 1 x 66.

## 4. Results and Discussion

For training purposes, the study used PyTorch in multi-GPU architecture. This dataset used 70-30 distribution. Seventy (70) percent of the dataset was used for training and the remaining thirty (30) percent was used for validation set. The training and validation were deployed using TITAN PC with 2 GPUs available. The list of classes and actual data was listed below.

```
# number of classes
nc: 17
```

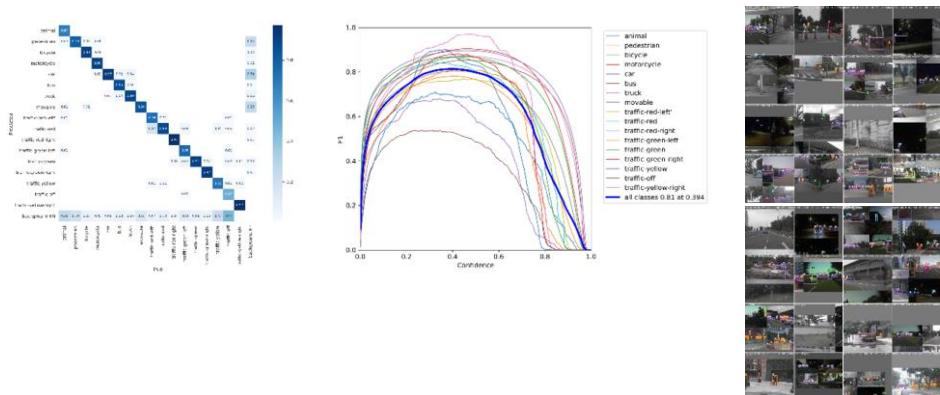
```
# class names
names: ['animal', 'pedestrian', 'bicycle', 'motorcycle', 'car', 'bus', 'truck', 'movable', 'traffic-red-left', 'traffic-red',
'traffic-red-right', 'traffic-green-left', 'traffic-green', 'traffic-green-right', 'traffic-yellow', 'traffic-off', 'traffic-yellow-right']
```

Using a pre-trained model at the beginning, we started to train the dataset using 2 GPUs.

After a series of iterations, the training stops when it meets certain criteria as follows:

- Sufficient iterations of at least 2000 for each class or at least less than the number of training images; and
- When the average loss no longer decreases (the lower, the better).

The study stopped at 25700 iterations and generated weights file indicated in the path written on the data file. Usually, YOLO generates the best and last weight files which show the highest mAp (see Figure 4). As seen in the Figures 4, shows f1 curve at 0.81 at 0.394 and all the classes stated in the confusion matrix shows acceptable kappa percentages.



**Figure 4.** F1 Curve and Confusion Matrix

This model was initially tested offline using recorded mp4 files and we found out some issues which arise because of the possible reasons (see Figure 5):

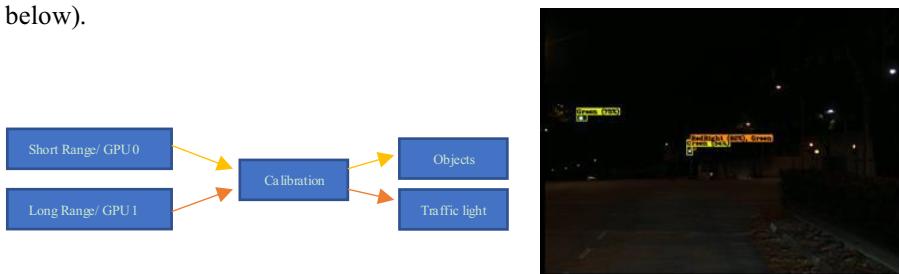
- imbalance number of samples for objects and traffic lights; and
- imbalance number of samples for daytime and night time.



**Figure 5.** Inconsistent Traffic Light Recognition

Based on the Figure 5, left image shows only 2 (two) traffic light signals and instead of four (4). Additionally, one of the two (2) detected traffic light signal should be green right (see the 3rd red circle), it only detected green. Another issue arise in the right image which shows the bursting of the lights during night time although the model is capable of some of it but the green right was not detected (see the red circle) since it is too close to the green traffic light signal.

To solve this, we deploy a calibration file and a long-range camera dedicated to the traffic light signals. The calibration file will adjust the parameters of the camera such as exposure, frames per second, and aperture for both daytime and night time (see the results below).

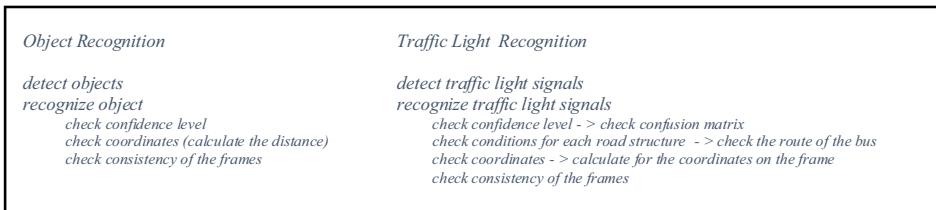


**Figure 6.** A more reliable and efficient traffic light recognition

Figure 6 shows the process we deploy to minimize the inconsistency that we encountered on the recognition. The two (2) cameras which deployed the same model has a dedicated GPU ID and a calibration file. The short range camera will be dedicated to react to objects only while the long range camera will be allotted to react to traffic light signals only.

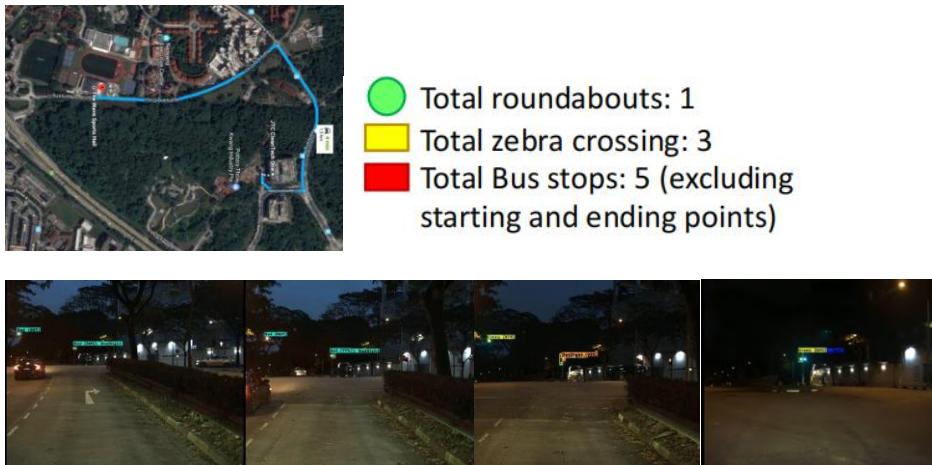
Since an acceptable model was generated with an mAP of ~90%, this model will then be integrated into the Autonomous Bus using ROS (Robotic Operating System). For this a vision\_darknet\_detect node was utilized. The model was implemented in a 50m Autonomous Bus within Nanyang Technological University route with a Nuovo PC 6108GC. The two (2) cameras as mentioned earlier, each of them is connected to a dedicated GPU using an ID. A GPU ID is dedicated for each camera during runtime. Once an object or traffic light signal was detected, the bus will depend its movement on the confidence level and the coordinates of the bounding box on the frame and the

consistency of the recognitions based on the number of frames. This follows the algorithm below:



**Figure 7.** Object and Traffic Light Recognition Algorithm

Since the bus route is very complex which comprises of one (1) roundabout, three (3) zebra crossings and five (5) bus stops and one (1) traffic light signals in the junction (see Figure 8 below). In the case of the junction, it must detect the greenright traffic signal from afar since the bus must be on the rightmost lane before approaching the junction.



**Figure 8.** Object and Traffic Light Recognition Algorithm

As the figure 8 shows from the leftmost image, the bus using a long-range camera and following the bus designated route, it detects and recognize the traffic light signal from afar and goes into the rightmost lane (the route is to turn right at the junction). While it is still at redlight, the bus will stop before the allotted stopping line and continues to detect the traffic light signal and when it satisfies the conditions set for confidence level, and the consistent detection of frames, once it recognizes greenlight, the bus will turn right at the junction.

## 5. Conclusion

This study was able to develop an acceptable model and deploy it on 50meter Autonomous Bus operating in day time and night time conditions. It was able to recognize objects in a real-time basis and generate a command to the bus. Though this study produced an acceptable and usable result, this study recommends to do sensor fusion incorporating lidar data together with the camera data for a higher safety consideration.

## References

- [1] Shaheen S, Finson R. Intelligent Transportation Systems. 2013 10.1016/B978-0-12-409548-9.01108-8.
- [2] Abbeel P, Ng AY. Apprenticeship learning via inverse reinforcement learning. In Proceedings of ICML, 2004.
- [3] Business Insider News: <http://www.businessinsider.com/companiesmaking-driverless-cars-by-2020-2017-1/#tesla-recently-made-a-bigmove-to-meet-its-goal-of-having-a-fully-self-driving-car-ready-by-2018-1> - accessed on March 24th, 2017.
- [4] Campbell K, Shia V, Bajcsy R. Decisions for autonomous vehicles: integrating sensors, communication and control. HiCoNS'14, April 15–17, 2014, Berlin, Germany. ACM 978-1-4503-2652-0/14/04.<http://dx.doi.org/10.1145/2566468.2576851>.
- [5] World Health Organization Website:<<http://www.who.int/mediacentre/factsheets/fs358/en>> - accessed on March 10th.
- [6] Liu W, Anguelov D et al. “ssd: Single shot multibox detector,” In ECCV, 2016.
- [7] Redmon J, Farhadi A “Yolov3: An incremental improvement,”[arXiv:1804.02767 \[cs.CV\]](https://arxiv.org/abs/1804.02767), 8 Apr 2018.
- [8] He K, Zhang X et al., “Spatial pyramid pooling in deep convolutional networks for visual recognition,” In CVPR, 2014
- [9] Girshick R “Fast r-cnn,” In ICCV, 2015.
- [10] Ren S, He K et al. Faster r-cnn: Towards real-time object detection with region proposal networks. In NIPS, 2017.
- [11] Everingham M, Eslami SMA, Van Gool L, Williams CKI, Winn J, Zisserman A. The pascal visual object classes challenge: A retrospective.” International J. Comp. Vision, 2015; 111(1): pp. 98–136.
- [12] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? the kitti vision benchmark suite, in CVPR, 2012
- [13] Lin TY, Maire M, Belongie S et al. Microsoft COCO: Common objects in context. in European Conference on Computer Vision, Oral, Jan. 1, 2014.

# FTR-Based Expert System for Power Generation Units

Naftaly MENN<sup>a,1</sup> and Boris CHUDNOVSKY<sup>a</sup>

*AMS – Advanced Measurement Systems Ltd., IEC- Israel Electric Company, Israel*

**Abstract.** The present expert system addresses the well-known problem existing on each large size boiler of a power generation utility - cleaning of fouling deposition created on the furnace walls and surfaces of heat exchangers. Continuous cleaning, which is a must, especially on large size units operated on fossil fuels, is very cumbersome and time-consuming procedure, mostly performed manually by technical staff sometime assisted by intelligence systems based on neural network. The expert system for cleaning suggested in the present study offers a new approach when the inference engine of the system is oriented to optimization of heat transfer inside the furnace. The target of optimization is maximization of the overall cleanliness factor of the furnace. On-going calculation of local heat transfer in different zones of the furnace is performed in real time by FTR devices positioned in different locations on the furnace wall and measuring continuously the Fouling Thickness and Reflectivity (FTR), along with local temperature of the flame. The system algorithm defines at each measurement cycle the optimal position where the next cleaning should be done and the corresponding signals are transferred by the system electronic circuitry to the controllers of the cleaning measures (the group of soot blowers to be activated). During the last two years our expert system has been successfully implemented on four coil-firing boilers of 500-600 MW on two power stations in Israel.

## 1. Introduction - Expert systems for Power Generation Utilities

It is well known that operation of power generation utility is a complicated task requiring on-going management and continuous attention of the engineering and technical staff of the station. Usually, utility is equipped with a great number of sensors and measuring devices allowing to follow after all stages of electricity generation, starting from fuel preparation and supply to the furnace, fuel combustion and heat transfer from the flame to the water converted in steam and delivery of the steam to the turbine where finally electric power is produced. Numerous results of measurements are transferred to the control room of the station (see, for example, description in Ref. [1]) where they are observed and analyzed by the operation technicians, which, according to the received information, decide when and how to interfere in order to provide correct and cost-effective electricity production process.

One of the most difficult for analysis is the situation in the furnace where the heat originated in the fuel combustion is transferred to the water/steam mixture. If a fuel mass is not burnt completely, what occurs inevitably when the utility is working with coal or biomass, unburnt particles are deposited on the water tubes surfaces (Fouling), creating

---

<sup>1</sup> Corresponding author, AMS – Advanced Measurement Systems Ltd., IEC- Israel Electric Company, Israel;  
E-mail: mnaftaly@gmail.com.

an essential thermal resistance to the heat flux. As a result, less heat is absorbed in the furnace tubes and, along with this, the furnace Flue Gas Exit Temperature (FEGT) is increasing which results in slagging, tubes damage, increasing of superheater and reheater spray and growing NOx emissions (see for instance, the graphs of calculations and measurements presented in Ref. [1,2]). To avoid or at least to reduce the negative effect of Fouling the furnace wall and the convection pass surfaces are periodically cleaned using the soot blowing devices which create the strong jets of turbulent air or steam breaking the Fouling deposits. A large size utility normally is equipped with 100-200 soot blowers activated several times every day, so the cleaning procedure is accompanied by significant expenses on water/ steam and electricity of soot blower's operation. Besides of that, soot blowing also causes significant erosion of the water tubes (see again indications in Ref [1, 2]). Obviously, it is desirable to decrease the number of cleaning cycles, but then the problem arises of when and where to activate the soot blowers. It is for this reason several **expert systems** (called sometime as **Intelligent soot blowing systems**) have been suggested (see Ref. [1-3]), all of them aiming to optimize the cumbersome procedure of boiler cleaning and to make it easier to implement them on power stations.

Two kinds of expert systems are presently in use. The first, described in details in Ref. [1], implement the algorithm based on some rules established in advance while analyzing relations between the boiler performance and activation of soot blowers in different zones of the boiler. The second, (see Ref. [3]), exploits the algorithms of neural network, providing the data base for learning is based on historically collected records while the boiler (of a proper design) was operated at some predefined conditions. Actually, during such learning experiments, similarly to the expert system from Ref. [1], the different groups of soot blowers were activated sequentially and FEGT and other heat transfer parameters have been estimated.

Obviously, both kinds of the expert systems facilitate decision making about cleaning activation, but the usefulness of the recommended procedure is limited because conditions during learning experiment may differ significantly from real situation on the operated facility where the type of fuel, the load of the boiler, the mode of operation and many other factors vary almost every day. Besides of that, as it is evident from description in Ref. [1-3], calculated parameters of heat transfer and the cleanliness are estimated while exploiting complicated models of heat transfer inside the furnace (like those presented in Ref. [4,5]) and do not use the directly measured parameters of the Fouling .This drawback can be avoided by using recently developed hardware - **FTR-Fouling Thickness and Reflectivity measurement system**, intended for direct non-contact real time measurement inside the operated furnace (see Ref.[6-8]). Usage of FTR enables to simplify calculation of dynamic heat transfer and renders new possibility for cleaning optimization. The corresponding new expert system is described in the following sections of this paper.

## 2. Our Approach

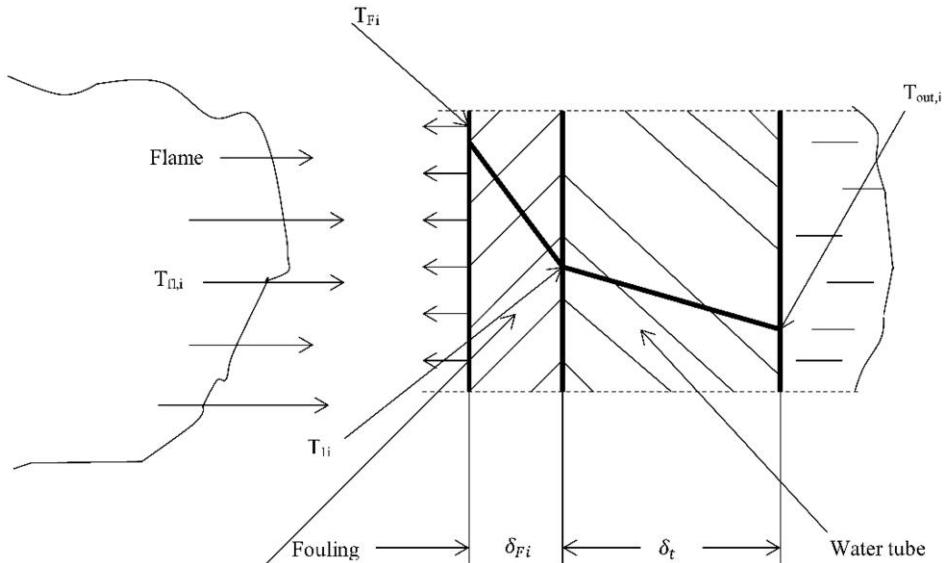
Keeping in mind that the goal of the expert system is the cleaning of the furnace, we address the cleaning as an optimization process and choose the Overall Cleaning Factor (CF) as a preferable parameter of optimization.

Let the furnace wall be divided to  $n$  zone cleaning of which is to be taken into account. Then. the Overall Cleanliness Factor CF can be expressed as a sum of  $n$  local

cleanliness factors  $CF_i$  each of which is defined as the ratio of the current heat flux  $Q_i$  transferred to the water/steam mixture to the heat flux  $Q_{0i}$  when the local fouling is completely removed (clean wall).

$$CF = \frac{1}{n} \frac{\sum_{i=1}^n Q_i}{\sum_{i=1}^n Q_{0i}} \quad (1)$$

In order to calculate the  $Q_i$  and  $Q_{0i}$  we take into account that at high temperatures developed in the furnace 95% of heat is transferred by radiation from the flame to the fouling layer and furthermore by conduction through the fouling and through the metal of tube to the water inside the tube. The general heat transfer model, described in Ref [4,5], can be significantly simplified for our consideration, as shown on Figure 1. Namely, the temperature distribution in both the Fouling layer and the wall tube is governed by the set of the following heat transfer equations:



**Figure 1.** Heat transfer model.

$$(1 - R_{Fi})\varepsilon\sigma(T_{fl,i}^4 - T_{Fi}^4) = \lambda_f \frac{T_{Fi} - T_{1i}}{\delta_{Fi}} = \lambda_t \frac{T_{1i} - T_{out}}{\delta_t} \quad (2)$$

Eq. (2) allows to get expression of  $T_{1i}$  through  $T_{Fi}$ :

$$T_{1i} = T_{Fi} \frac{\lambda_f/\delta_F}{\frac{\lambda_F}{\delta_{Fi}} + \frac{\lambda_t}{\delta_t}} + T_{out,i} \frac{\lambda_t/\delta_t}{\frac{\lambda_F}{\delta_{Fi}} + \frac{\lambda_t}{\delta_t}} \quad (3)$$

Eq. (3) in the Eq. (2) results in the following non-linear equation with regards to  $T_{Fi}$ :

$$(1 - R_{Fi})\varepsilon\sigma(T_{fl,i}^4 - T_{Fi}^4) - T_{Fi} \left[ \frac{\lambda_F}{\delta_{Fi}} - \frac{\left( \frac{\lambda_F}{\delta_{Fi}} \right)^2}{\frac{\lambda_F}{\delta_{Fi}} + \frac{\lambda_t}{\delta_t}} \right] + T_{out,i} \frac{\frac{\lambda_F * \lambda_t}{\delta_{Fi} * \delta_t}}{\frac{\lambda_F}{\delta_{Fi}} + \frac{\lambda_t}{\delta_t}} = 0 \quad (4)$$

which obviously can be solved numerically. Getting  $T_{Fi}$  allows to calculate the local heat flux  $Q_i$  in zone  $i$  when its reflectivity is  $R_{Fi}$  and its fouling thickness achieves the value  $\delta_{Fi}$ :

$$Q_i = \lambda_t \frac{T_{Fi} - T_{out,i}}{\delta_t} \quad (5)$$

Now we take into account that just after cleaning  $\delta_{Fi} \rightarrow 0$  and  $T_{Fi} \rightarrow T_{1i}$ .

Then, instead of Eq. (4) we receive

$$(1 - R_{Fi0})\varepsilon\sigma(T_{fl,i}^4 - T_{1i,0}^4) = \lambda_t \frac{T_{1i,0} - T_{out,i}}{\delta_t} \quad (6)$$

Solving (6) with regard of  $T_{1i,0}$ , again numerically, we get the values of temperature on the surface of the clean tube which enables to calculate the heat fluxes through the cleaned furnace:

$$Q_{0i} = \lambda_t \frac{T_{1i,0} - T_{out,i}}{\delta_t} \quad (7)$$

Going back to Eq. (1) we rewrite it as follows:

$$CF = \frac{1}{n} \frac{\sum_1^n CF_i * Q_{0i} * W_i}{\sum_1^n Q_{0i} * W_i} \quad (8)$$

where the local Cleanliness Factors  $CF_i$  are introduced as well as the correction factors  $W_i$  which could be used if additional parameters affecting the impact of the zone  $i$  on the Overall Cleanliness Factor should be taken into consideration.

Let at the moment  $t_1$  the situation on the furnace is described by the local Cleanliness factors

$$CF_i^{(t_1)} = \frac{Q_i^{(t_1)}}{Q_{0i}} \quad (9)$$

and suppose in the moment  $t_2$  the cleaning will be done in zone  $k$ . Then obviously  $Q_k = Q_{0k}$ , so that from Eq. (1) for the moment  $t_2$  we receive

$$\begin{aligned} CF^{(t_2)} &= \frac{\sum_1^{k-1} CF_i * Q_{0i} * W_i + Q_{0k} * W_k + \sum_{k+1}^n CF_i * Q_{0i} * W_i}{n \sum_1^n Q_{0i} * W_i} \\ &= \frac{\sum_1^{k-1} Q_i * W_i + Q_{0k} * W_k + \sum_{k+1}^n Q_i * W_i}{n \sum_1^n Q_i * W_i} = CF_{new}^{(t_2)} \end{aligned} \quad (10)$$

The last expression shows what would be the Cleanliness Factor of the furnace if at  $t_2$  only area number  $k$  will be cleaned. Repeating calculations for all  $k$  from 1 to  $n$  and comparing the results we can find the zone which corresponds to the maximum new Cleanliness Factor, in other words the cleaning in this area will improve the heat transfer more than cleaning of any other zone and therefore can be defined as optimal (of course for this very moment of time):

$$k_{opt} : \max \{CF_{new}^{(t_2)}\} \quad (11)$$

The time difference ( $t_2 - t_1$ ) is the duration of cleaning cycle after which the optimization procedure should be activated again and so on, meaning that the results of (11) renders the full sequence of locations recommended for cleaning. This recommended sequence can be either used for manual operation or implemented automatically, as described in the following sections.

### 3. Data base and the Hardware

As it can be seen from the previous description, the suggested expert system requires continuously performed calculation of heat transfer inside the furnace. Parameters needed for calculation vary from zone to zone (the fouling thickness  $\delta_{FI}$  and reflectivity  $R_{FI}$ , and the local temperature of the flame  $T_{FLi}$ ) and also are time dependent. Therefore, the data base for this expert system cannot be collected in advance, but have to be measured directly and in real time.

In order to carry out these direct measurements during normal operation of the boiler a special technique and a special hardware called FTR-Fouling Thickness and Reflectivity system (see Ref. [ 6,7]), have been exploited. It comprises a number of specially designed devices positioned in different zones of the furnace wall, each having a moving probe which is introduced into the furnace through a small opening, 6mm by 12 mm, on the membrane between the wall tubes. The probe is extended into the furnace for a short time of measurement (up to 2 sec) and then is withdrawn outside. Each FTR device includes the light source (a small laser diode) creating a light spot on the surface of Fouling which is imaged on the video detector. Position of the spot image on the detector renders information about the Fouling thickness whereas intensity of light reflected back to the probe and incident on the detector gives information about the local reflectivity of the Fouling before and after the cleaning.

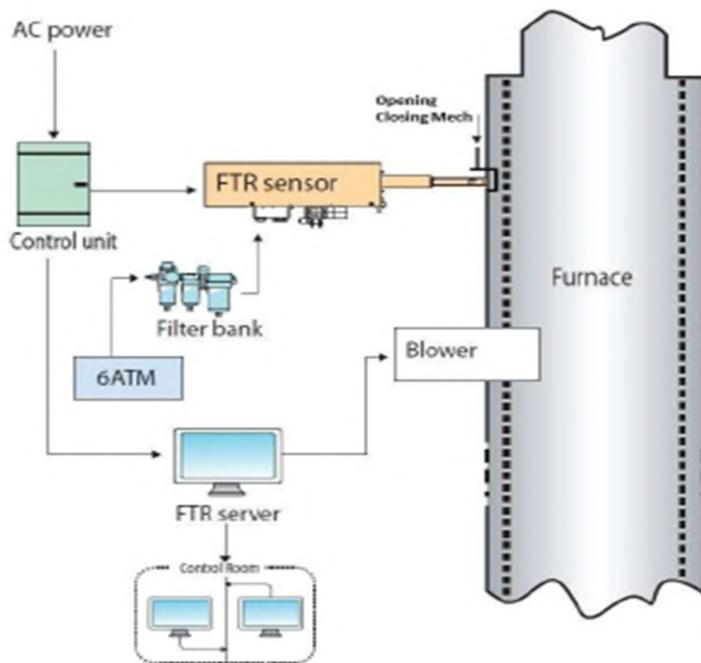
Recently an advanced FTR device was suggested (see Ref. [8]) which, along with the Fouling sensor, comprises also the flame temperature sensor. Using the principle of two wavelengths optical pyrometer, this sensor enables to measure the local temperature of the flame and the flame local apparent emissivity.

A full expert system includes a number of FTR devices, each with a processing unit, all of them being interfaced to the central computer where the algorithm of optimization described in section 2 is continuously run and analyzed at every cleaning cycle (usually each 30 min).

Figure 2 demonstrates schematics of the FTR device and its accessories, all are operated by the suggested expert system.

### 4. Implementation

The expert system described above has been implemented on two power stations in Israel. Two coal firing boilers on each station have been equipped with this expert systems- 575 MW tangential firing boilers and 550 MW opposite firing boilers. On both stations optimal cleaning was first realized manually, while decision making was assisted by the system algorithm presented in section 2 above, and later it was operated fully automatically when the calculation results from the system server were transferred to the electronic controllers activating the soot blowers of the furnace.



**Figure 2.** Schematics of FTR hardware on the furnace wall.

Automatic operation of the expert system on these two power stations is continuously working during the last two years. In the course of this period of time it was realized that several additional rules should be added to the main algorithm of the system allowing to stop automatic cleaning when regime of the boiler is changed drastically or in order to avoid approaching to some critical situations. Execution of the rules is either examined by the system server or transferred to the system processor as a chain of messages sent from the station control room.

Figure 3a and 3b demonstrates how effective is the cleaning optimization governed by the automatic expert system. On Figure 3a are presented the measurement results (he Fouling thickness and hemispherical reflectivity) before activation of the FTR system, the red lines indicate when the soot blowers were activated according to the procedure existing on the power plant. Figure 3b shows how drastically the program of cleaning was changed using optimal cleaning by FTR system – the frequency of soot blower's activation was reduced more than two times without significant growth of the Fouling deposition thickness and reflectivity and even getting improved Cleanliness Factor (CF initial value was 0.7 while the final value achieved 0.85).

## The FTR system generates unique data:

Sootblowing sequence – without FTR, with Heat Flux Sensors, Cleanliness Factor = 0.7

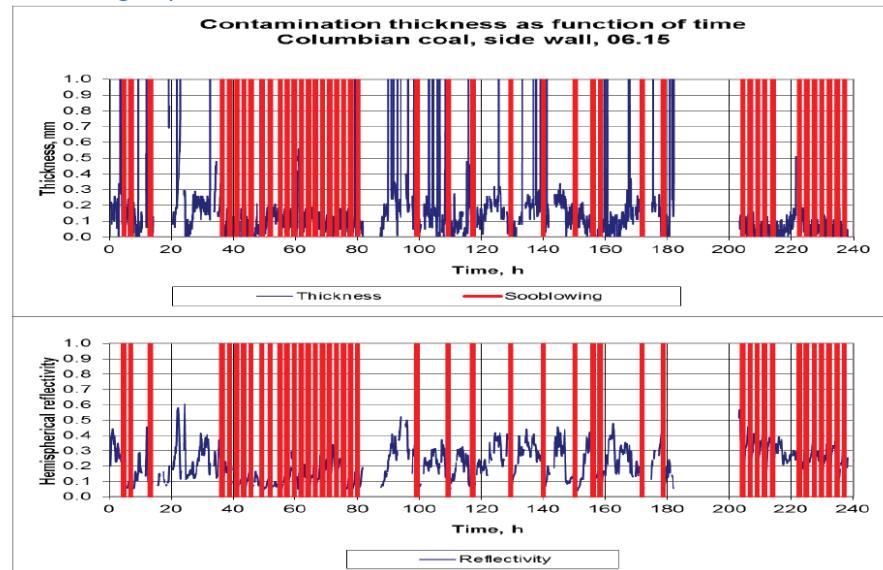


Figure 3a. Measurement results without FTR.

## Soot blowing sequence Controlled by the FTR System - Cleanliness Factor = 0.85

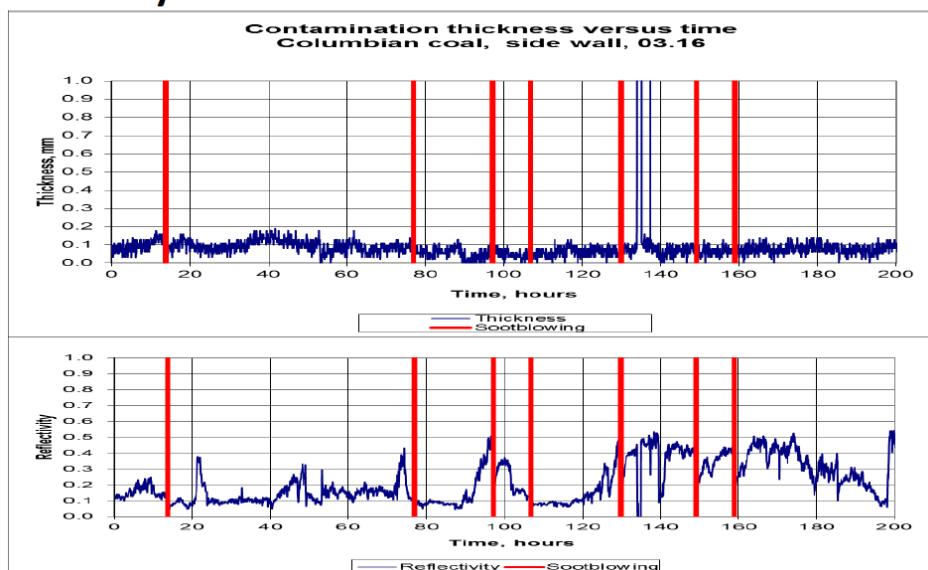


Figure 3b. Measurement results with activated FTR.

## 5. Summary

A new expert system allowing automatic optimization of the cleaning procedure in the furnace of a coal firing boilers is suggested. The system database is created in real time while measuring the major heat transfer parameters - the flame temperature and emissivity along with fouling deposition thickness and reflectivity. The measurements are carried out by specially designed FTR devices based on electro-optical arrangements allowing non-contact direct measurement of major parameters of heat transfer inside the furnace. The main algorithm of the expert system performs optimization procedure at each cleaning cycle with regard to Overall Cleanliness Factor of the furnace.

The system was implemented on two power station in Israel, where optimal cleaning enables to get significant reduction of expenses on operation of the soot blowers and improves performance of the boilers (higher CF).

## List of Symbols

CF – Cleanliness Factor

$Q$  – heat flux,  $W/m^2$

T- absolute temperature in  $^{\circ}K$

$\varepsilon$  – apparent emissivity of the flame

$\delta$  – thicknes of the fouling layer or a tube wall,  $m$

R = reflectivity of fouling deposition

$\lambda$  – thermal conductivity,  $W/(m \cdot K)$

$\sigma$  – Stefan – Boltzman constant

W – factor of relative influence on heat transfer

n – a number of zones at the furnace wall, taken into consideration

## Subscript

fl -- flame

F - fouling

t = water tube

i - current value in zone i

$0i$  - value in zone  $i$  just after cleaning

## References

- [1] Custer J, Fuller TA. Performance improvements through intelligent sootblowing optimization at longview power. *Power Engin.* 2017; 121(2).
- [2] Piboontum SJ, Swift SM, Conrad RS. Boiler performance improvement due to intelligent soot blowing utilizing real-time boiler modeling UP Boilers. Presented to Electric Power 2005, April 5-7; Chicago, Illinois, USA.
- [3] Shah S, Bergemann C, Breeding C. Implementation of intelligent soot blowing. Clyde Bergemann, Inc. Electric Power 2004; Baltimore, Maryland, March 30 – April 1, 2004.
- [4] Zhang YG, Li QH, Zhou H. Heat transfer calculation in Furnaces. In book: Theory and calculation of heat transfer in Furnaces, Academic Press, December, 2016.

- [5] Makarov AN. Calculations of heat transfer in the furnaces of steam boilers according to the laws of radiation of gas volumes. Chapter 6 in: Heat transfer- Models, Methods and Applications. IntechOpen, 2018, p.111-140.
- [6] Naftaly M, Joseph K. A method and apparatus for measurement of thickness and reflectivity of fouling deposits on tubes in a Pulverized coal -firing Furnace. US patent No. 9,709,384, July 18, 2017.
- [7] Menn N, Chudnovsky B. New technology for monitoring fouling deposition in coal-firing boilers. Intern. J. Electr. Heat Generation VGB PowerTech. 2016; 6: p.65.
- [8] Menn N, Krimerman J. Measurement and Monitoring of the local flame temperature, flame apparent emissivity, fouling thickness and fouling reflectivity in the furnace of a coal and/or biomass firing boiler. Provisional Patent Application, 63/259,642, 08/02/2021, USA.

# Context-Sensitive Control of Adaptation: Self-Modeling Networks for Human Mental Processes Using Mental Models

Raj BHALWANKAR <sup>a</sup>, Laila VAN MENTS <sup>b</sup> and Jan TREUR<sup>a,1</sup>

<sup>a</sup>*Social AI Group, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands*

<sup>b</sup>*AutoLeadStar, Jerusalem, Israel*

**Abstract.** Within their mental and social processes, humans often learn, adapt and apply specific mental models of processes in the world or other persons, as a kind of blueprints. In this paper, it is discussed how analysis of this provides useful inspiration for the development of new computational approaches from a Machine Learning and Network-Oriented Modeling perspective. Three main elements are: applying the mental model by internal simulation, developing and revising a mental model by some form of adaptation, and exerting control over this adaptation in a context-sensitive manner. This concept of controlled adaptation relates to the Plasticity Versus Stability Conundrum from neuroscience. The presented analysis has led to a three-level computational architecture for controlled adaptation. It is discussed and illustrated by examples of applications how this three-level computational architecture can be specified based on a self-modeling network and used to model controlled learning and adaptation processes based on mental models in a context-sensitive manner.

**Keywords.** Mental model, Adaptive network, Control of adaptation, Self-modeling network

## 1. Introduction

Historically, analysis of human neural, mental and social processes often has provided inspiration for interesting new methods and techniques in AI and Machine Learning. In this paper, the focus is on how in their mental and social processes, humans often learn, adapt and apply specific mental models as a kind of blueprints, schemas or maps. Although a vast majority of the large amounts of publications on mental models in multiple disciplines is informal and not computational, it is shown how analysis of them still can be used as inspiration for the development of new computational approaches from an AI and Machine Learning perspective. Three elements came out of this analysis: (1) applying a mental model by internal (mental) simulation, (2) to develop and maintain a mental model, adaptation of it takes place (which usually involves learning, extinction or forgetting, and revision), and (3) in a context-sensitive manner, control is exerted over this adaptation. Here in particular (3) is an interesting topic. In a wider neuroscientific context, for example, in [1], this topic is discussed in relation to what is called the Plasticity Versus Stability Conundrum and to the notion of metaplasticity; e.g., [2]. It has

<sup>1</sup> Corresponding Author: Jan Treur, Social AI Group, Vrije Universiteit Amsterdam, De Boelelaan 1111, 1081HV Amsterdam, the Netherlands; Email: j.treur@vu.nl.

been found that in the brain plasticity is not constant and various neural mechanisms have been discovered by which the extent of plasticity varies by being controlled in a context-sensitive manner; [1, 2, 3, 4]. Also, within an AI-context, in machine learning examples of this conundrum and controlled adaptation to address it are known, such as the (decreasing) temperature parameter in simulated annealing and the sensitive balancing between exploration and exploitation in reinforcement learning, also called the explore-exploit dilemma [5, 6, 7]. This is explained in [7] as follows.

‘When you go to your favourite restaurant, do you always order the same thing, or do you try something new? Sticking with an old favourite ensures a good meal, but if you are willing to explore you might discover something better. This simple conundrum, deciding between something you know and something you do not, is commonly referred to as the exploration–exploitation dilemma.’

This quote also illustrates that on the one hand decision making based on known mental models can be very efficient (like navigating based on a well-known map), but on the other hand this may prevent someone from learning even better decisions (like exploring still unknown territory). The analysis of how humans use mental models in their mental processes has led to a three-level computational architecture for context-sensitive controlled adaptation where at the base level internal simulation based on a mental model takes place, at the next level adaptation of this mental model and at the second-next level context-sensitive control over this adaptation. In this paper, it is pointed out and illustrated by examples how this three-level computational architecture can be modeled as a self-modeling network: a network in which parts of the network’s own structure characteristics are represented by nodes within the same network.

## 2. How Humans Use Mental Models

In their mental processes, humans use mental models in three different ways: for internal simulation, for learning, and for (self-)control; these uses are explained as follows.

**Simulation: Mental Models Simulate.** Mental models are often used for internal simulation. This can take place in different forms, for example, visualisation or flashbacks (in sport or in PTSD). In his often cited book [8], Kenneth Craik describes a mental model as a *small-scale model* that is carried by an organism within its head:

‘If the organism carries a “small-scale model” of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations ...’ ([8], p. 61)  
 ‘...it is a physical working model which works in the same way as the process it parallels...’ ([8], p. 51).

Internal simulation is assumed to be based on causal relations for world processes; e.g., [9, 10, 11, 12]. For example, in [13], the functioning of a car in interaction with its driver is simulated based on a person’s mental model, and in [14] a flashback experience of a course of events is addressed as internal simulation based on a mental model of this course of events.

**Adaptation: Mental Models Adapt.** Mental models are adapted often (e.g., learning, revising, forgetting); for example:

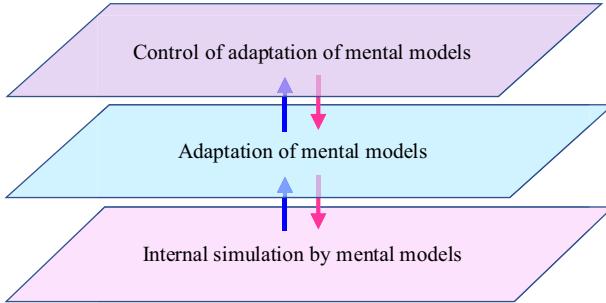
‘Model-based learning is a dynamic, recursive process of learning by building mental models. It incorporates the formation, testing, and subsequent reinforcement, revision, or rejection of mental models of some phenomenon.’ [15]

*Observational learning* concerns observation of others or of oneself while ‘learning by doing’ or ‘learning by discovery’. Learners may see someone perform a behavior and

then start to imitate it; e.g., [16, 17]. Mirror neurons are a basis for learning by observation; e.g., [18, 19, 20]. On the other hand, *instructional learning* is a useful addition as only learning based on observation may be not very efficient; e.g., [21].

**Control: Mental Models Respond to Control.** For learning mental models, *control* is exerted; see also [22, 23, 24]. Such control is also referred to as metacognition; e.g., [25, 26, 27, 28]. Also, the term self-regulation is used [28].

**Overall architecture.** Based on the processes for mental models summarized above, a cognitive architecture for handling mental models has been designed covering the three types of processes in an integrated manner as depicted in Figure 1.



**Figure 1.** A cognitive architecture for mental model handling.

Here, mental models are addressed through multiple representations: they can be viewed from three levels of representation according to the three planes depicted in Figure 1. At the base level depicted by the lower (pink) plane, a mental model, which essentially is considered to be a relational structure, is represented by its nodes and the relations between the nodes. For internal simulation, the nodes have activation levels that vary over time: based on the relations these activation levels affect each other over time. Next, at the adaptation level depicted in Figure 1 by the middle (blue) plane, it is represented how the mental model relations change over time by some adaptation specification. Finally, at the top level depicted by the upper (purple) plane in Figure 1 it is indicated how the adaptation at the middle level is controlled. For more details of this architecture and its application, see [29].

### 3. Modeling by Self-Modeling Networks

The Network-Oriented Modelling approach based on adaptive temporal-causal networks as self-modeling networks from [30, 31] is a suitable modeling approach to represent in an adaptive and controlled manner the causal relations and the way they can be processed to generate mental processes. Using this approach, adaptive networks of multiple orders can be modelled relatively easily. Network nodes  $X$  have state values indicated by real numbers  $X(t)$  that vary over time  $t$ ; nodes are also called states. The characteristics defining a network are:

- **Connectivity characteristics:** connections from states  $X$  to  $Y$ , having *connection weights*  $\omega_{X,Y}$  specifying their strengths
- **Aggregation characteristics:** each state  $Y$  has a *combination function*  $\mathbf{c}_Y$  that specifies how impact from all incoming connections on  $Y$  is aggregated. Based on

a list of basic combination functions  $\text{bcf}_i$  (each with some parameters) provided by an available library, such a combination function can be specified by weights  $\gamma_i$  and parameters  $\pi_{ij}$  for these basic combination functions  $\text{bcf}_i$

- **Timing characteristics:** each state  $Y$  has a *speed factor*  $\eta_Y$  specifying how fast  $Y$  changes

The numerical representation created by the available dedicated software environment is based on the following equations based on the above network characteristics (where  $X_1, \dots, X_k$  are the states from which state  $Y$  gets incoming connections):

$$\mathbf{impact}_{X,Y}(t) = \omega_{X,Y} X(t) \quad (1)$$

$$\mathbf{aggimpact}_Y(t) = \mathbf{c}_Y(\mathbf{impact}_{X_1,Y}(t), \dots, \mathbf{impact}_{X_k,Y}(t)) = \mathbf{c}_Y(\omega_{X_1,Y} X_1(t), \dots, \omega_{X_k,Y} X_k(t)) \quad (2)$$

$$\begin{aligned} Y(t+\Delta t) &= Y(t) + \eta_Y [\mathbf{aggimpact}_Y(t) - Y(t)] \Delta t \\ &= Y(t) + \eta_Y [\mathbf{c}_Y(\omega_{X_1,Y} X_1(t), \dots, \omega_{X_k,Y} X_k(t)) - Y(t)] \Delta t \end{aligned} \quad (3)$$

A computational network engine developed within this software environment based on the generic equations (3) takes care for the processing of all network states thereby using their connections and other network characteristics.

**Self-modeling networks to model adaptivity and control.** To design network models that are adaptive, the concept *self-modeling network* (also called reified network) introduced in [30, 31] has turned out to be very useful. A self-modeling network is obtained if for some of the network characteristics  $\omega, \gamma, \pi, \eta$  as introduced above, network states are added to the network that represent their value. For example, for a connection weight  $\omega_{X,Y}$  an additional state  $\mathbf{W}_{X,Y}$  (called a *first-order self-model state*) is added to the network that represents the value of this weight  $\omega_{X,Y}$  and is indeed used as value for that weight in the processing. For such an additional network  $\mathbf{W}$ -state, also additional network characteristics are added to get an adequate embedding in the obtained self-modeling network. For example, for Hebbian learning [32] a specific combination function  $\mathbf{hebb}_\mu(V_1, V_2, W)$  from the library can be used involving a persistence parameter  $\mu$ :

$$\mathbf{hebb}_\mu(V_1, V_2, W) = V_1 V_2 (1-W) + \mu W \quad (4)$$

Here,  $V_1, V_2$  are activation levels of the connected states and  $W$  is the activation level of the self-model state  $\mathbf{W}_{X,Y}$  for the connection weight. As a simpler alternative, sometimes the function

$$\mathbf{smin}_\lambda(V_1, V_2) = \min(V_1, V_2)/\lambda \quad (5)$$

is used for Hebbian learning, where  $\lambda$  is a scaling factor; e.g., see [33].

As an example at a next (*second-order self-model*) level, for the combination function  $\mathbf{hebb}_\mu(V_1, V_1, W)$  of a self-model state  $\mathbf{W}_{X,Y}$ , the persistence parameter  $\mu$  can be represented by another self-model state  $\mathbf{Mw}_{X,Y}$ . The latter network state is a *second-order self-model state* as it represents a network characteristic related to (first-order) self-model state  $\mathbf{W}_{X,Y}$ . Another example of a second-order self-model state is a state  $\mathbf{Hw}_{X,Y}$  that represents the speed factor (learning rate)  $\eta$  of (first-order) self-model state  $\mathbf{W}_{X,Y}$ . These

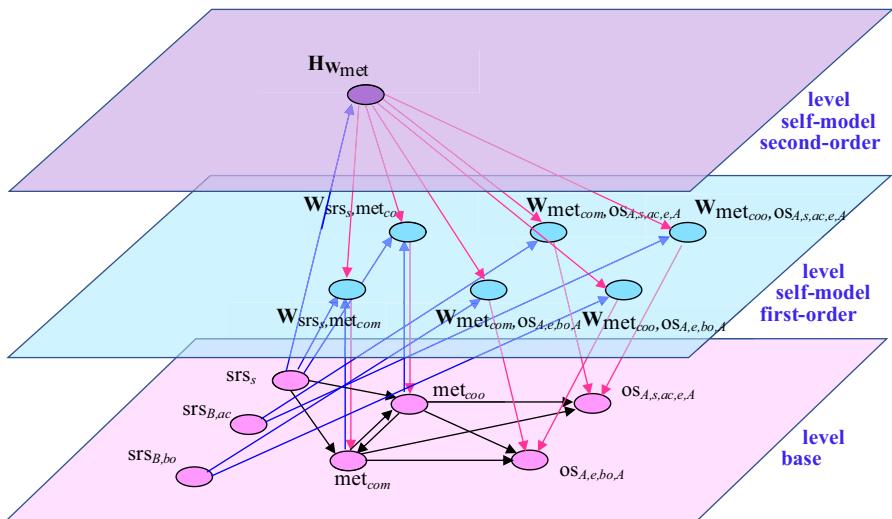
two types of self-model states can be used in an adaptive network model to control the learning of mental models: if the activation level of  $\mathbf{Mw}_{xy}$  gets lower, more extinction takes place which occurs when mental models are forgotten. Moreover, decreasing activation values of  $\mathbf{Hw}_{x,y}$  model make the learning slow down, and a value 0 causes a complete freezing of the learning (no plasticity).

These first- and second-order self-model states directly relate to the three levels in the architecture depicted in Figure 1. Within the lowest (pink) plane the base network states are depicted, within the middle (blue) plane the first-order self-model states (such as states  $\mathbf{W}_{x,y}$ ), and within the upper (purple) plane the second-order self-model states (such as states  $\mathbf{Mw}_{x,y}$  and  $\mathbf{Hw}_{x,y}$ ).

#### 4. Examples of Self-Modeling Network Models for Mental Models

In this section, as illustrations two examples of self-modeling network models for mental models are briefly discussed. They are both instantiations of the generic three-level architecture discussed in Section 2.

**Using metaphors as mental models.** This example (described in more detail in [34]) addresses the use of metaphors as mental models to support joint decision making; see Figure 2 for a part of this network model. At the base levels two different metaphor options are modeled: cooperative metaphor state  $\text{met}_{coo}$  and competitive metaphor state  $\text{met}_{com}$ ; see the lower (pink) plane in Figure 2. The incoming and outgoing connections for these metaphor states are adaptive, modeled by the first-order self-model  $\mathbf{W}$ -states; see the middle (blue) plane. The control of the adaptation of all these  $\mathbf{W}$ -states is modeled by the second-order self-model  $\mathbf{Hw}$ -state; see the upper (purple) plane.

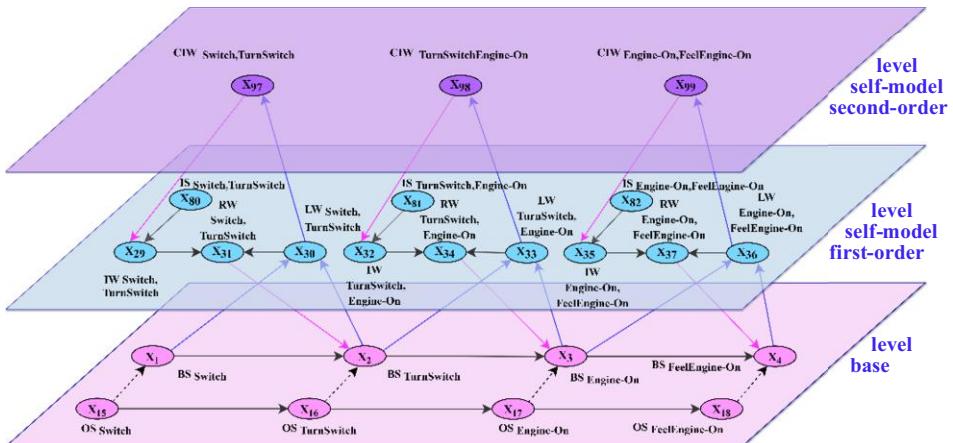


**Figure 2** Connectivity of a part of an adaptive network model for mental models for metaphors with their activation and effect connections at the base level, their learning at the first-order self-model level and the control of the learning at the second-order self-model level.

The two  $\mathbf{W}$ -states for the incoming connections for the metaphor states determine when the metaphor should be activated. These connections are learnt by Hebbian

learning, which has as effect that the more often they are used in a certain context (represented by stimulus representation  $srs_s$ ), the stronger their activations become when that context occurs. The four **W**-states for the outgoing connections from the metaphor states determine the effect of the metaphor states on the self-owner states; they are adapted based on whether the other person tends to go for action  $ac$ , and whether the other person has a good feeling  $bo$  about it (states  $srs_{B,ac}$  and  $srs_{B,bo}$ ). For each of these six **W**-states, its adaptation is controlled through the second-order self-model **Hw**-state which represents the learning rate of the related **W**-state. This form of second-order adaptation is based on the second-order adaptation principle ‘adaptation accelerates with increasing stimulus exposure’ [4].

**Learner-controlled learning.** The second example (described in more detail in [35]) addresses how a learner controls at what times instruction by a teacher takes place; see Figure 3 for a representative part of this model. Here the three **RW**-states in the middle (blue) plane represent the weights of the connections in the mental model (that is learned) depicted at the (pink) base level plane. For example, state **RW<sub>Switch,TurnSwitch</sub>** within the first-order self-model plane, also indicated by  $X_{31}$ , represents the weight of the connection in the base plane from state **BS<sub>Switch</sub>** to state **BS<sub>TurnSwitch</sub>** (the arrow from  $X_1$  to  $X_2$ ).

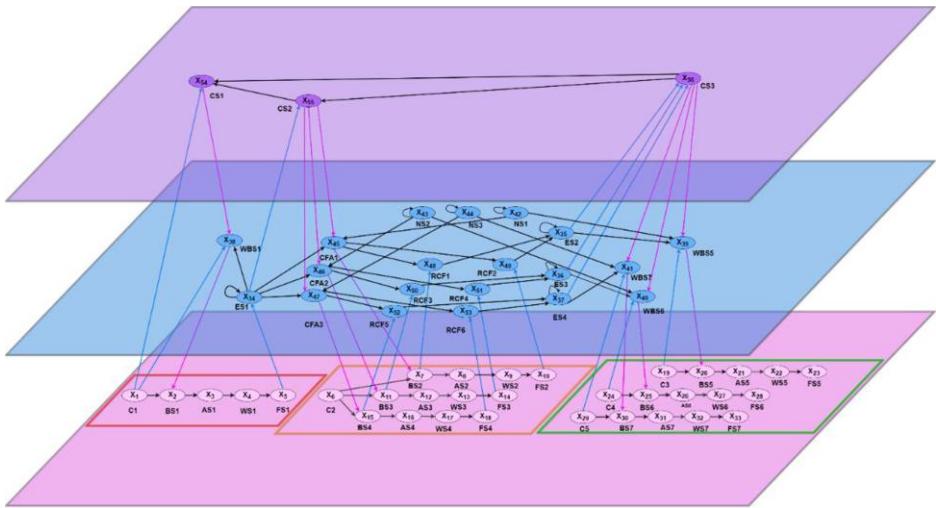


**Figure 3.** Connectivity of a part of an adaptive network model for mental models for learner control of a teacher’s instruction.

These **RW**-states are based on input from observational learning (modeled by the corresponding **LW**-states) and instructional learning (modeled by the corresponding **IW**-states). For example, state **RW<sub>Switch,TurnSwitch</sub>** has incoming connections from **LW<sub>Switch,TurnSwitch</sub>** and **IW<sub>Switch,TurnSwitch</sub>**. The **LW**-states are adaptive based on Hebbian learning; to this end they use combination function (4) from Section 3. For this scenario, they are not (explicitly) controlled. In contrast, for the **IW**-states, their incoming connections from the **IS**-states (representing the communication channel from the information source **IS**-state, which is the teacher) are controlled. This control is modeled by the second-order self-model **CTW**-states at the upper (purple) level; these are also denoted as **Ww**-states. For example, state **CIW<sub>Switch,TurnSwitch</sub>** in the (purple) second-order self-model plane controls the connection from **IS<sub>Switch,TurnSwitch</sub>** to **IW<sub>Switch,TurnSwitch</sub>**, which

models the communication from instructor to learner about the connection from BS<sub>Switch</sub> to state BS<sub>TurnSwitch</sub>. Such a control state CIW<sub>Switch,TurnSwitch</sub> is activated once the corresponding LW-state LW<sub>Switch,TurnSwitch</sub> has a sufficient action level: the learner has (partially) learnt it based on observation and asks the teacher for confirmation. So, these CIW-states determine when the communication channels have to be opened for the instructional learning.

**Learning by counterfactual thinking.** A third case to illustrate the approach is adopted from [36]; it addresses learning by counterfactual thinking; see Figure 4. Here after a negatively evaluated experience modelled by the red box in the base plane, the person considers internal simulation of mental models for alternative action options and their expected results (see the orange box in the base plane). The counterfactual status of this is modelled by the process modelled in the middle blue plane and that process is controlled by the upper level CS-states in the purple plane. See [36] for more details.



**Figure 4.** Connectivity of the controlled adaptive network model for counterfactual thinking.

## 5. Options for Control of Adaptation using Self-Modelling Networks

In Section 4, some examples of adaptation control for mental models were discussed, for two cases within the context of self-modelling networks used as a vehicle. In this section, a more systematic overview is discussed of different ways in which control over mental models and their adaptation can be exerted, thereby again making use of the network structure provided by self-modelling networks. In a controlled adaptive network model for mental models based on a self-modelling network, adaptation is modelled by a first-order self-model. There are a number of network characteristics involved in the first-order self-model states used for the adaptation. In particular, by systematically going through all possible network characteristics, the following examples of network characteristics for adaptation to be controlled can be distinguished and are illustrated by various examples:

- **Control by adaptive connectivity characteristics of first-order self-model states**
  - Adaptive connections of the causal pathways to the self-model states and their weights  $\omega$ ; for example:
    - Choosing a mental model to be applied. For example, a decision to use a specific metaphor-based mental model as in the model in Section 4 taken from [34] or a decision to use a geometric mental model to support learning of an arithmetic mental model, as described in [37]
    - Opening a communication channel from an information source to enable instructional learning of a mental model (decision to ask), as in the model in Section 4 taken from [35] and in the model described in [38]
    - Opening an observation channel to enable observational learning of a mental model (decision to observe), as in the model described in [38]
  - Adaptive connections of the causal pathways from the self-model states to other states and their weights  $\omega$ ; for example:
    - Modelling the effects of a chosen metaphor as in the model in Section 4 taken from [34]
- **Control by adaptive aggregation characteristics of first-order self-model states**
  - Adaptive choice of combination function; for example:
    - For Hebbian learning of mental model connections a weighted average of  $\text{hebb}_\mu(V_1, V_2, W)$  and  $\text{smin}_\lambda(V_1, V_2)$ , with adaptive weights  $\gamma_1$  and  $\gamma_2$ .
  - Adaptive parameters of chosen combination functions; for example:
    - Adaptive values for the persistence factor  $\mu$  of  $\text{hebb}_\mu(\cdot)$  as in the self-modelling network model for shared mental models described in [38] or for the scaling factor  $\lambda$  of  $\text{smin}_\lambda(V_1, V_2)$ .
- **Control by adaptive timing characteristics of first-order self-model states**
  - Adaptive adaptation speed (learning rate)  $\eta$ ; for example:
    - Addressing the Plasticity Versus Stability conundrum [1] based on some context factors indicating when plasticity is needed fully and when plasticity should be limited or frozen.
    - Accelerating adaptation speed upon increased stimulus exposure [4], for example as applied in the example model in Section 3 taken from [34]

In Table 1 an overview is given of a large number of applications of such self-models to model human processes based on mental models as brought together in book [29].

Recall from Section 3 how at some level self-model states can be introduced to represent network characteristics from a lower level. The naming is as follows: self-model state  $\mathbf{W}_{X,Y}$  represents connection weight  $\omega_{X,Y}$  from the lower level,  $\mathbf{H}_Y$  represents speed factor  $\eta_Y$  from the lower level, and so on. Such self-model states can be generally called **W-states** or **H-states**, for example. By iterating this, for example, second-order self-model state  $\mathbf{Hw}_{X,Y}$  represents the adaptation speed of first-order self-model state  $\mathbf{W}_{X,Y}$  and second-order self-model state  $\mathbf{Mw}_{X,Y}$  represents the persistence parameter  $\mu_{\mathbf{W}_{X,Y}}$  of first-order self-model state  $\mathbf{W}_{X,Y}$ . Similarly, second-order self-model state  $\mathbf{Wz}_{Z,W_{X,Y}}$  represents the weight  $\omega_{Z,W_{X,Y}}$  of the connection from some state  $Z$  to state  $\mathbf{W}_{X,Y}$ . Such second-order self-model states can be generally called **Hw-states**, **Mw-states**, or **Ww-states**. Note that in most cells in Table 1 further references are included, but for those cells where no references are included, in general this means that these options are yet to be addressed in detail.

**Table 1** Overview of different types of controlled adaptation for example mental models as collected in [29]

Learning by	Control of adaptation via connectivity	Control of adaptation via aggregation	Control of adaptation via timing
Observation and monitoring	Controlled learning by observation for a mental model $\mathbf{W}$ -state for bonding via a $\mathbf{W}_w$ -state ([29], Ch 13; formation of a mental model of another person; see also [33])	Hebbian mental model learning $\mathbf{W}$ -state persistence control via an $\mathbf{M}_w$ -state ([29], Ch 14; controlled forgetting of a mental model relation; see also [39])	Hebbian mental model learning $\mathbf{W}$ -state speed control via an $\mathbf{H}_w$ -state: adaptation accelerates with increasing exposure ([29], Ch 5, Ch 7, Ch 10, Ch 11, Ch 12; e.g., learning mental models for flashback experiences [14], analysis and support tasks [40], metaphors [34], God-models [41], self- and other-models [33])
Excitability adaptation	Incoming connection for an adaptive mental model excitability $\mathbf{T}$ -state control via a $\mathbf{W}_T$ -state ([29], Ch 7; learning excitability [42, 43] of a mental model's states; see also [40])	Excitability mental model learning $\mathbf{T}$ -state aggregation control, for example through adaptive (steepness $\sigma$ and threshold $\tau$ ) parameters of a logistic combination function used for the $\mathbf{T}$ -state represented by $\mathbf{S}_T$ - and $\mathbf{T}_T$ -states	Excitability mental model learning $\mathbf{T}$ -state speed control via an $\mathbf{H}_T$ -state ([29], Ch 7; learning excitability [42, 43] of a mental model's states; see also [40])
Communication	Learner-controlled instructional learning of a mental model $\mathbf{W}$ -state via a $\mathbf{W}_w$ -state ([29], Ch 9; opening a communication channel with the instructor by asking; see also [35]) Controlled learning by communication for a mental model $\mathbf{W}$ -state for bonding via a $\mathbf{W}_w$ -state ([29], Ch 13 opening a communication channel with the other person by asking; see also [38])	Learner-controlled instructional learning of a mental model $\mathbf{W}$ -state via a $\mathbf{T}_w$ -state for excitability [42, 43] of the $\mathbf{W}$ -state (opening a communication channel with the instructor by more sensitive listening)	Learner-controlled instructional learning of a mental model $\mathbf{W}$ -state via an $\mathbf{H}_w$ -state (controlling the timing of a communication channel with the instructor)
Other mental models	Controlled connection $\mathbf{W}$ -state for counterfactual activation of a mental model via a $\mathbf{W}_w$ -state ([29], Ch 6; see also [36])	Controlled connection $\mathbf{W}$ -state for counterfactual activation of a mental model via a $\mathbf{T}_w$ -state addressing excitability [42, 43] of the $\mathbf{W}$ -state	Controlled inter mental model exchange connection $\mathbf{W}$ -state via $\mathbf{H}_w$ -state ([29], Ch 4; exchange from arithmetic mental model to geometric mental model; see also [37])
Contextual factors	Controlled connection $\mathbf{W}$ -state for activation of mental model via a $\mathbf{W}_w$ -state ([29], Ch 6; activation of a mental model for possible future action; see also [36])	Controlled connection $\mathbf{W}$ -state for activation of mental model via a $\mathbf{T}_w$ -state addressing excitability of the $\mathbf{W}$ -state based on contextual factors	Controlled adaptive mental model effect connection $\mathbf{W}$ -state via an $\mathbf{H}_w$ -state ([29], Ch 10; adapting the own choices based on the context given by the other person; see also [34])

## 6. Discussion

This paper addressed how in mental and social processes, humans often apply specific mental models and learn and adapt them in a controlled manner. It was explored how analysis of such processes may provide useful inspiration for the development of computational approaches from an AI, computational intelligence and machine learning

perspective. It was discussed how controlled adaptation relates to the Plasticity Versus Stability Conundrum in neuroscience [1]. From the analysis a three-level computational architecture for controlled adaptation was obtained. It was discussed how this three-level computational architecture can be modeled as a self-modeling network [30, 31], illustrated by a number of examples.

The work discussed here can be considered as part of the area of human-like machine learning. Also, within machine learning in general there are a few approaches that do (explicitly) address control of adaptation. A simple example is Simulated Annealing, where the temperature parameter can be considered an adaptive adaptation control parameter. Another example is the Explore-Exploit Dilemma in reinforcement learning, where the challenge is to find criteria on the context that can be used to decide in a context-sensitive manner between explore or exploit; e.g., [5, 6, 7]. Two limitations of the current state of the presented work that may be identified are:

- The approach has only been explored yet for deterministic models. In principle, probabilistic models are possible when combination functions are added to the library with some probabilistic or random effects, but this still has to be explored in more detail
- Analysis of the wider domain of human-like approaches to control of adaptation, in particular for mental model adaptation as discussed here, may provide inspiration for further development of machine learning methods. This is also an area that has not yet been explored in more detail.

It may be noted from this paper that for human-like controlled adaptation much input and inspiration was obtained from the area of neuroscience; for example, in [1, 2, 3, 4] and many other papers on metaplasticity several second-order adaptation principles for context-sensitive control of adaptation can be found. This is in strong contrast with the area of social science. Literature on adaptive social relations or adaptive social networks is available including computational models for it, such as [44, 45], but although adaptation control is certainly also in this area a very relevant factor, approaches addressing the control of adaptation seems to be lacking in such literature, where only just a few more or less incidental exceptions can be found such as [6, 30, 34, 38, 39, 46, 47] and [31], Ch 6. Much progress would become possible for social science by explicitly addressing this issue more systematically and on a much larger scale like it was already done within neuroscience.

## References

- [1] Sjöström PJ, Ranz EA, Roth A, Häusser M. Dendritic excitability and synaptic plasticity. *Physiol Rev* 2008; 88: 769–840.
- [2] Abraham WC, Bear MF. Metaplasticity: the plasticity of synaptic plasticity. *Trends in Neuroscience* 1996; 19(4):126-130.
- [3] Magerl W, Hansen N, Treede RD, Klein T. The human pain system exhibits higher-order plasticity (metaplasticity). *Neurobiology of Learning and Memory* 2018; 154:112-20.
- [4] Robinson BL, Harper NS, McAlpine D. Meta-adaptation in the auditory midbrain under cortical influence. *Nat. Commun.* 2016; 7: article 13442.
- [5] Holland JH. Adaptation in natural and artificial systems. Ann Arbor, MI: University of Michigan Press; 1975.
- [6] March JG. Exploration and exploitation in organizational learning. *Organization Science* 1991; 2: 71-87.
- [7] Wilson RC, Geana A, White JM, Ludvig E, Cohen JD. Humans use directed and random exploration to solve the explore-exploit dilemma. *J. Experimental Psychol: General* 2014; 143(6): 2074–2081.
- [8] Craik KJW. The nature of explanation. Cambridge, MA: University Press; 1943.

- [9] Damasio AR. *Descartes error: emotion, reason and the human brain*. London: Vintage Books; 1994.
- [10] Goldman AI. *Simulating minds: the philosophy, psychology, and neuroscience of mindreading*. Oxford University Press, New York; 2006.
- [11] Hsslöw G. Conscious thought as simulation of behaviour and perception. *Trends Cogn. Sci.* 2002; 6(2002), 242-247.
- [12] Hsslöw G. The current status of the simulation theory of cognition. *Brain Res.* 2012; 1428: 71-79.
- [13] Bhalwankar R, Treur J. Modeling the development of internal mental models by an adaptive network model. In: Proc. of the 11th Annual International Conference on Brain-Inspired Cognitive Architectures for AI, BICA\*AI'20. Procedia Computer Science, Elsevier Publishers; 2021.
- [14] Van Ments L, Treur J. A higher-order adaptive network model to simulate development of and recovery from PTSD. In: Paszynski M, Kranzlmüller D, Krzhizhanovskaya VV, Dongarra JJ, Sloot PMA (eds) *Computational Science – Proc. ICCS 2021. Lecture Notes in Computer Science*, vol. 12743. Springer Nature, Cham; 2021. p. 154-166.
- [15] Buckley BC. Interactive multimedia and model-based learning in biology. *Intern. J. Sci. Education* 2000; 22(9): 895-935.
- [16] Benbassat J. Role modeling in medical education: the importance of a reflective imitation. *Academic Medicine* 89(4); 550-554: 2014.
- [17] Yi MY, Davis FD. Developing and validating an observational learning model of computer software training and skill acquisition. *Information Syst. Res.* 14(2); 146-169: 2003.
- [18] Hurley S. The shared circuits model (SCM): How control, mirroring, and simulation can enable imitation, deliberation, and mindreading. *Behavioral Brain Sci.* 2008; 31(1), 1-22.
- [19] Rizzolatti G, Craighero L. The mirror-neuron system. *Annu. Rev. Neurosci.* 2004; 27: 169-192.
- [20] Van Gog T, Paas F, Marcus N, Ayres P, Sweller J. The mirror neuron system and observational learning: Implications for the effectiveness of dynamic visualizations. *Educ. Psychol. Rev.* 2009; 21(1): 21-30.
- [21] Seel NM. Mental models in learning situations. In: *Advances in Psychology*, vol. 138. Amsterdam: North-Holland; 2006. p. 85-107.
- [22] Gibbons J, Gray M. An integrated and experience- based approach to social work education: The Newcastle model. *Social Work Education* 2002; 21: 529-549.
- [23] Hogan KE, Pressley ME. *Scaffolding student learning: Instructional approaches and issues*. Brookline Books; 1997.
- [24] Kozma RB. Learning with media. *Review of educational research* 1991; 61(2): 179-211.
- [25] Darling-Hammond L, Austin K, Cheung M, Martin D. Thinking about thinking: Metacognition. In *The Learning Classroom: Theory into Practice*. Stanford University School of Education 2008. p. 157-172.
- [26] Mahdavi M. An overview: metacognition in education. *Intern. J. Multidisciplinary Current Res.* 2014; 2 :529-535.
- [27] Koriat A. Metacognition and consciousness. In: Zelazo PD, Moscovitch M, Thompson E (eds.). *Cambridge Handbook of Consciousness*. New York: Cambridge University Press. 2007.
- [28] Pintrich PR. The role of goal orientation in self-regulated learning. In: Boekaerts M, Pintrich P, Zeidner M (eds.). *Handbook of self-regulation research and applications*. Orlando, FL: Academic Press. 2000, p. 451-502.
- [29] Treur J, Van Ments L (eds.). *Mental Models and their Dynamics, Adaptation and Control: A Self-modeling Network Modeling Approach*. Springer Nature, Cham. 2022.
- [30] Treur J. Modeling higher-order adaptivity of a network by multilevel network reification. *Network Science* 2020; 8: S110-44.
- [31] Treur J. Network-oriented modeling for adaptive networks: designing higher-order adaptive biological, mental, and social network models. Cham, Switzerland: Springer Nature Publishing; 2020. 412 p.
- [32] Hebb DO. *The organization of behavior: A neuropsychological theory*. New York: John Wiley and Sons; 1949. 335 p.
- [33] Hermans A, Muhammed S, Treur J. An Adaptive network model for attachment theory. In: Paszynski M, Kranzlmüller D, Krzhizhanovskaya VV, Dongarra JJ, Sloot PMA (eds) *Computational Science – Proc. ICCS 2021. Lecture Notes Comp. Sci.*, vol 12744. Springer Nature, Cham. 2021. p. 462-475.
- [34] Van Ments L, Treur J. Modeling adaptive cooperative and competitive metaphors as mental models for joint decision making. *Cognitive Systems Research* 2021; 69: 67-82.  
<https://doi.org/10.1016/j.cogsys.2021.06.002>.
- [35] Bhalwankar R, Treur J. A second-order adaptive network model for learner-controlled mental model learning processes. In: Proc. of the 9th International Conference on Complex Networks and their Applications, vol. 2. *Studies in Computational Intelligence*, vol. 944. Springer Nature Switzerland AG; 2021. p. 245-259. Extended version to appear in PLOS One.
- [36] Bhalwankar R, Treur J. 'If Only I Would Have Done that...': A Controlled Adaptive Network Model for Learning by Counterfactual Thinking. In: Proceedings of the 17th International Conference on Artificial

- Intelligence Applications and Innovations, AIAI'21 Advances in Information and Communication Technology, vol. 627. Springer Nature; 2021. p. 3-16.
- [37] Treur J. An adaptive network model covering metacognition to control adaptation for multiple mental models. *Cognitive Syst. Res.* 2021; 67: 18-27.
- [38] Treur J. Modeling controlled social network adaptation using mental models. Proc. of the 7<sup>th</sup> International Congress on Information and Communication Technology, ICICT'21. Advances in Intelligent Systems and Computing, Springer Nature, in press. 2021.
- [39] Van Ments L, Treur J, Klein J, Roelofsma PHMP. A second-order adaptive network model for shared mental models in hospital teamwork. In: Nguyen NT, et al. (eds.), Proc. of the 13th International Conference on Computational Collective Intelligence, ICCCI'21. Lecture Notes in AI, Springer Nature; 2021
- [40] Treur J. Self-modeling networks using adaptive internal mental models for cognitive analysis and support processes. In: Proc. of the 9th International Conference on Complex Networks and their Applications, vol. 2. Studies in Computational Intelligence, vol. 944. Springer Nature Switzerland AG. 2021, p. 260-274. Extended version: A Self-Modeling Network Model Addressing Controlled Adaptive Mental Models for Analysis and Support Processes. Complex Systems, 2021, in press.
- [41] Van Ments L, Treur J, Roelofsma PHMP. How empathic is your god: an adaptive network model for formation and use of a mental god-model and its effect on human empathy. In: Treur J, Van Ments L (eds.), Mental Models and their Dynamics, Adaptation and Control: A Self-modeling Network Modeling Approach. Springer Nature, Cham; 2022.
- [42] Debanne D, Inglebert Y, Russier M. Plasticity of intrinsic neuronal excitability. *Current Opinion in Neurobiology* 2019; 54: 73-82.
- [43] Chandra N, Barkai E. A non-synaptic mechanism of complex learning: modulation of intrinsic neuronal excitability. *Neurobiology of Learning and Memory* 2018; 154: 30-36.
- [44] McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. *Annu. Rev. Sociol.* 2001; 27: 415-44.
- [45] Treur J. Mathematical analysis of the emergence of communities based on coevolution of social contagion and bonding by homophily. *Appl. Network Sci.* 2019; 4: article 1.
- [46] Carley KM. Inhibiting Adaptation. In: Proceedings of the 2002 Command and Control Research and Technology Symposium. Monterey, CA: Naval Postgraduate School; 2002. p. 1-10.
- [47] Carley KM. Destabilization of covert networks. *Comput. Math Organiz. Theor.* 2006; 12: 51-66.

# Domain Adaptation for Document Image Binarization via Domain Classification

Carlos GARRIDO-MUNOZ<sup>a,1</sup>, Adrián SÁNCHEZ-HERNÁNDEZ<sup>a</sup>,

Francisco J. CASTELLANOS<sup>a</sup> and Jorge CALVO-ZARAGOZA<sup>a</sup>

<sup>a</sup>Dept. of Software and Computing Systems, University of Alicante, Alicante, Spain

**Abstract.** Binarization represents a key role in many document image analysis workflows. The current state of the art considers the use of supervised learning, and specifically deep neural networks. However, it is very difficult for the same model to work successfully in a number of document styles, since the set of potential domains is very heterogeneous. We study a multi-source domain adaptation strategy for binarization. Within this scenario, we look into a novel hypothesis where a specialized binarization model must be selected to be used over a target domain, instead of a single model that tries to generalize across multiple domains. The problem then boils down to, given several specialized models and a new target set, deciding which model to use. We propose here a simple way to address this question by using a domain classifier, that estimates which of the source models must be considered to binarize the new target domain. Our experiments on several datasets, including different text styles and music scores, show that our initial hypothesis is quite promising, yet the way to deal with the decision of which model to use still shows great room for improvement.

**Keywords.** Document Image Binarization, Deep Neural Networks, Unsupervised Domain Adaptation, Domain Classifier

## 1. Introduction

Binarization represents a key role in many Document Image Analysis (DIA) workflows [1,2], as it helps to reduce the complexity of the image by highlighting the relevant information (i.e., the ink). This process also enables the use of specific procedures involving morphological operators, connected-component search, or histogram analysis, among others. Given its importance in the field of DIA, there is a vast amount of existing literature concerning document image binarization, including surveys and reviews [3].

A straightforward procedure for binarization is that of thresholding, in which all pixels under a certain value are set to 0, being set to 1 otherwise. For example, Otsu's algorithm automatically estimates a global threshold for a given input image [4]. In contrast to global thresholding, there also exist algorithms that compute a different threshold for each pixel depending on its local neighborhood, such as Niblack's [5], Sauvola's [6] and that proposed by Wolf et al. [7]. More complex algorithms for document image binarization have also been proposed, such as those by Gatos et al. [8], Su et al. [9] or Howe

---

<sup>1</sup>Corresponding Author: Carlos Garrido-Munoz, Dept. of Software and Computing Systems, University of Alicante, Alicante, Spain; E-mail: cgm156@alu.ua.es

[10], all of which are based on different image processing workflows comprising several steps.

Supervised learning has been also considered for document image binarization. A classification-based approach consists in querying every pixel of the image, performing a feature extraction, and then using a learned model to predict its category. Within this formulation, both classical Multi-Layer Perceptron [11] and Convolutional Neural Networks [12] have been studied.

More recently, image-to-image models based on Fully Convolutional Neural Networks (FCN), which were first proposed for semantic segmentation [13], have been applied to document image binarization [14]. The FCN takes an input grayscale or color image and directly provides the probability of each pixel to be foreground or background in just one step. According to one of the latest Competition on Document Image Binarization [15]—a common benchmark for this task—the use of FCN can be considered the current state of the art, as most of the best methods are based on this idea. Therefore, we will use this kind of neural network as the backbone of our methodology.

It is convenient to emphasize, however, that it is very difficult for the same model to work successfully in a number of document styles, since the set of potential domains is very heterogeneous. Previous work demonstrated that the use of a model over a different type of manuscript to that for which it was trained noticeably decreased the performance [16].

The common idea to deal with the situation mentioned above is to train a single model with datasets of different types, with the hope that it will generalize to unknown domains successfully. However, our hypothesis is that a specialized model could better deal with the binarization of a target domain. The open question, therefore, is how to decide which of these specialized models should be used or to what extent one should trust their predictions, conditioned to the characteristics of the target set to be binarized. We study here a straightforward way to address this question by using a domain classifier, which determines the suitability of each specialized binarization model for the new target domain.

Our experiments on several datasets, including different text styles and music scores, show that our initial hypothesis is quite promising, yet the way to deal with the decision of which model to use still shows great room for improvement.

The rest of the work is structured as follows: in Section 2, we thoroughly describe our methodology; Section 3 specifies our experimental setup are given; in Section 4, we report the results attained, along with analysis and discussion; in Section 5 a qualitative evaluation is discussed; and finally, in Section 6, we present the conclusions, including some potential ideas for future work.

## 2. Methodology

Image binarization is a function  $\mathcal{B} : [0, 255]^{(h \times w \times c)} \rightarrow \{0, 1\}^{(h \times w)}$  that converts an image with a defined height  $h$ , width  $w$ , and  $c$  channels into its binary version.

As mentioned above, the state of the art considers image-to-image FCN to approximate  $\mathcal{B}$ . Typically, these methods assume that the distribution of the data to be binarized is similar to that used for training, which is not very useful in many real cases. We here propose a strategy to deal with this situation.

Our methodology assumes an unsupervised multi-source domain adaptation scenario, where there is a collection of training sets  $\{\mathcal{S}_1, \mathcal{S}_2, \dots\}$ , each representing a source labeled domain with pairs of document images and their perfectly binarized versions, being  $\mathcal{X}^i = [0, 255]^{h_s^i \times w_s^i \times c}$  the  $i$ -th image from the training sets and  $\mathcal{Y}^i = \{0, 1\}^{h_s^i \times w_s^i}$  its corresponding ground-truth data. The goal, however, is to provide a binarization strategy that works successfully over a completely unknown target set, which is only presented at the test time.

In our strategy, a specialized binarization model is first trained for each of the source domains. Our hypothesis is that a new target set should be binarized with the best model among those of the previous step. We propose a domain classifier for this decision that knows how to distinguish between all these source sets. This classifier is later used to classify a test sample, and we then provide a binarization accordingly through a specific decision mechanism.

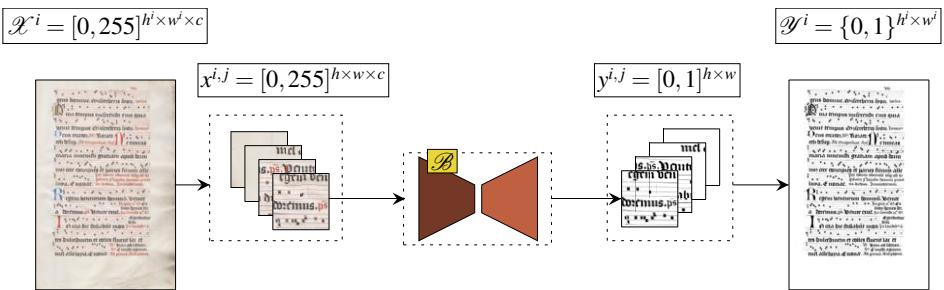
Below we describe thoroughly the three elements involved in our methodology: the binarization model, the domain classifier, and the decision mechanism.

### 2.1. Selectional Auto-Encoders

Although our strategy is independent of the underlying binarization model considered, as long as it is based on machine learning, in this work we shall implement the Selectional Auto-Encoder (SAE) model, proposed in the work of Calvo-Zaragoza and Gallego [14].

An SAE model is trained to perform a function such that  $b : \mathbb{R}^{(h \times w \times c)} \rightarrow [0, 1]^{(h \times w)}$ . That is, it learns a score map over a  $h \times w \times c$  image that preserves the input shape. The score value (or neural activation) of each pixel depends on whether the pixel belongs to the foreground or the background.

This type of architecture is typically implemented as an FCN, and so the prediction can be done through successive convolutions and sampling operations, without any fully-connected layer. The last layer consists of a set of neurons that predict a value in the range of  $[0, 1]$ , depending on the selectional level of the corresponding input pixel. A graphical illustration of this configuration is shown in Figure 1.



**Figure 1.** Schema of an SAE model in document image binarization. Note that  $\mathcal{X}^i$  is the  $i$ -th image from the source set while training, or the target set otherwise.

The weights of the SAE are learned through a training process with the aim at binarizing document images. The training stage consists of providing examples of images and their corresponding binarized ground-truth data. Since an SAE is a type of feed-forward network, the training process can be carried out by conventional means [17].

Once the SAE has been properly trained, an image can be parsed, after which a score level is assigned to each input pixel. In practice, the network barely outputs either 0 or 1 but an intermediate value. Therefore, a thresholding process is still necessary to convert the obtained neural activations into actual binary values. Since the network already takes into account the context of each pixel in its internal operation, a single global threshold is sufficient. In practice, unless otherwise stated, a threshold value of 0.5 can be assumed.

Furthermore, it might happen that the size of the input document is higher than the input layer of the SAE. In such a case, we simply split the input document into pieces of the size expected by the network and parse them independently. Then, to reconstruct the original image, we assemble the independent pieces, without further processing.

## 2.2. Domain Classifier

At this point, we have a way to train a specialized binarization model for each source domain ( $\mathcal{B}_i$ ). However, when a new test sample must be binarized, we do not know which of them should be used. To address this issue, our strategy includes a domain classifier that is responsible for estimating which of these specialized models is more appropriate at each case.

To achieve this goal, we use a very straightforward premise: we train a classifier that predicts which of the source domains a certain sample belongs to. This is simple to carry out, as we can create ground-truth data for domain classification by taking samples of the images from each source domain. Then, the model can be trained to categorize them as samples of the domain they were taken.

Among the options available to build this classifier, we consider Convolutional Neural Networks, because of their successful performance in image processing tasks. Note that a neural network does not fully categorize the input but produces a weight  $\hat{w} \in [0, 1]$  (output activation) for each possible category—domains, in this case. The weights represent the probability of the input sample belonging to each possible domain. As we describe in the next section, we shall propose alternatives for taking advantage of  $\hat{w}$  to select how the final binarization is performed.

## 2.3. Combination Mechanism

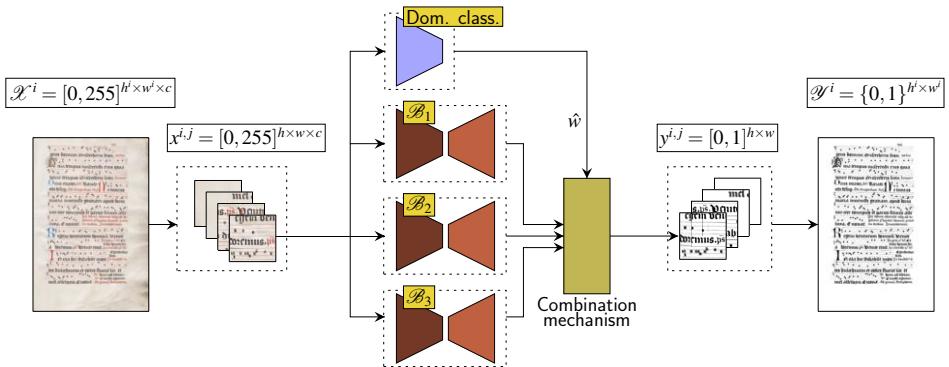
The only thing left to complete our strategy is to decide how to combine the binarization results provided by each of the specialized SAE models and the probabilities  $\hat{w}$  provided by the domain classifier.

From the received information, we must produce a single binarization result, which represents the actual output of our domain adaptation approach. We consider two alternatives for the combination of information:

- Class-based selection: the domain with the highest probability according to the domain classifier is selected for binarization.
- Weighted combination: the final score of each pixel is computed as a linear combination of the prediction by each specialized SAE, weighted by their  $\hat{w}$ .

The first option follows a *winner-takes-it-all* approach, whereas the second one ensembles the predictions of each specialized SAE according to their estimated confidence for the test data.

For the sake of clarity, we provide a general overview of our strategy in Figure 2, including the different elements and the data flows.



**Figure 2.** Framework schema proposed for multi-source domain adaptation for document image binarization.

### 3. Experimental setup

This section details the experimentation carried out to evaluate the proposed approach. First, the corpora and metrics considered are described, and second, the experimental set-up is detailed, with the architecture of the neural networks considered and their hyper-parameterization.

#### 3.1. Corpora

In order to assess our proposal, documents of different types have been considered, including text documents and music score images. Figure 3 includes representative examples from each and Table 1 shows the details of each one.

- DIB: set of images of handwritten Latin text documents from the Document Image Binarization Contest [18], annually held from 2009. We combined the collections from 2009 to 2016 as a single corpus for our experiments.
- PHI: collection of scanned images of Persian manuscripts from the Persian Heritage Image Binarization Competition [19] as a different text domain.
- SAL and EIN: two collections of high-resolution images of scanned documents that contain lyrics and music scores in neumatic notation. Specifically, those images from Salzinnes Antiphonal (CDM-Hsmu 2149.14)<sup>2</sup> and Einsiedeln, Stiftsbibliothek, Codex 611(89)<sup>3</sup>, respectively.

<sup>2</sup><https://cantus.simssa.ca/manuscript/133/>

<sup>3</sup><http://www.e-codices.unifr.ch/en/sbe/0611/>



**Figure 3.** Examples of document patches from the corpora considered.

**Table 1.** Details of the corpora. The columns represent the corpus, the number of images, the average resolution of the images, and finally, the average amount of ink pixels with respect to the entire images, respectively.

Corpus	Pages	Size	Ink
<i>Text documents</i>			
DIB	86	659 × 1560 px.	7.2%
PHI	15	1022 × 1158 px.	9.2%
<i>Music documents</i>			
SAL	10	5100 × 3200 px.	19.2%
EIN	10	5550 × 3650 px.	20.0%

For the experimentation, the images from the corpora have been configured with the 5-fold cross validation technique with three partitions, for training, validating and testing with 60%, 20% and 20% of the whole collections, respectively.

### 3.2. Metrics

The binarization issue is a two-class problem in which the pixels are classified as foreground or background. However, as seen in Table 1, given the imbalanced scenario in this context, the assessment requires to use metrics that do not bias towards the majority class—background in this case. For this, we consider the calculation of the *F-measure* ( $F_1$ ), which is mathematically defined as

$$F_1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}, \quad (1)$$

where TP represents the *True Positives* or correctly classified pixels, FP is *False Positives* or type I errors and FN stands for the *False Negatives* or type II errors. Note, that in our context, we assume the foreground as the positive class.

In addition to computing this metric for the binarization predictions, we also consider the multi-class version of this metric to assess the isolated performance of the domain classifier proposed.

### 3.3. Hyper-parameterization

Concerning the SAE architecture, it is divided into two parts: an encoder, composed of consecutive blocks of convolutional layers and down-sampling operators, and a decoder,

which includes blocks of convolutional layers and up-sampling operators, as many up-sampling operations as there are down-sampling ones in the encoder stage. For our experiments, we define an encoder block as two consecutive convolutional layers with the same number of filters in both and followed by a max-pooling operator of  $2 \times 2$  px. In addition, after all these encoder blocks, two additional convolution layers are included with 128 filters each. Concerning the decoder block, it is defined as an up-sampling operator of  $2 \times 2$  px. and two convolutional layers with a symmetric number of filters with respect to the encoder blocks. We particularly considered two blocks for both the encoder and decoder and a residual connection between the second blocks of each one after the convolution layers.

It should be noted that the number of filters is increased with the number of blocks in the encoder stage, and decreasing in the decoder one, in a factor of 2, with an initial number of 32 filters. We considered using a kernel shape of  $3 \times 3$  px. for these layers and a Rectified Linear Unit (ReLU) activation for each convolution. Finally, another convolutional operation with 32 filters and a kernel size of 15 is performed to eventually finish in a  $1 \times 1$  convolutional layer of one filter with Sigmoid activation to obtain the score result for each pixel of the input image.

In addition to the SAE model, we also need to define the domain classifier used for our proposal, which is the one responsible for determining which of the known domains—sources—most closely resembles to the new unlabeled domain. The architecture considered for this purpose contains three sequential blocks of a convolution layer, a max-pooling operator of  $2 \times 2$  px. and a dropout with 0.2 of probability. The first block uses 32 filters with a kernel of  $5 \times 5$  px. whereas the remaining two convolutions contain 64 filters with the same kernel. After these blocks, a flatten operation is performed followed by a fully-connected layer with 64 neurons. Finally, another fully-connected layer is used, with as many neuron as known domains, with a softmax activation. The rest of the convolutions use the ReLU activation function.

As aforementioned in Section 2, both the SAE model and the domain classifier do not process the full image at once, but they deal with patches in which the images are split. After informal experiments, we consider a size of  $256 \times 256$  px. for these patches. Moreover, the models are trained for 300 epochs maximum, with a batch size of 32 and checking the models in each epoch with the validation partitions to keep the best configuration. It is worth highlight that the learning process is carried out by means of the Adam optimizer [20] in the case of the binarization model, and stochastic gradient descent [21] for the domain classifier. Note that the test partition is not used at all for the training process.

It is worth mentioning that, in our experiments, we have considered RGB images (that is,  $c = 3$ ). However, any other color/grayscale space could be considered as long as the model is trained and used consistently.

### 3.4. Competing Approaches

For the sake of clarity in the analysis of the results, in this section, we here describe and denote each scenario considered in our experiments.

First, we considered two possible baseline experiments: the case in which only a single source domain is used and the one in which a single model is trained with all the source training sets. Henceforth, we denote these baseline cases as *Single-domain* and *Multi-domain*, respectively.

Another scenario to be studied is that in which our proposals are considered. As aforementioned, given a new image, a domain classifier recommends which of all the available specialized models should be used to binarize the image. The recommendation of the classifier is used in two different modes (see Section 2.3): the first one would be the case in which it decides the binarization model to be used according to the classification, and the second one, that obtains the probability of that image to belong to each source domains for then weighting the probabilities obtained for each binarization model. These two proposals are referred to in the rest of the paper as *Class-based selection* and *Weighted combination*, respectively.

Finally, since our approach is based on the decision of using a specific model or the weighted ensemble, we also include a study to analyze the upper bound that could be achieved with our hypothesis. This is artificially assessed by selecting always the best binarization model for each test patch. This experiment is referred to as *Ideal-patch*. Another ideal case to be analyzed is the case in which the domain classifier selects the best specialized model given a full domain, so that the same binarization model is applied for all the patches. We denote this experiment as *Ideal-global*. Note that these experiments are only used as a reference to measure the potential of our premise, but they do not represent real results.

## 4. Results

In this section, we present the final results achieved by our approach. Moreover, we compare them with the ones obtained by baseline methods, described in Section 3.4. Our results are reported in Table 2.

As can be observed, comprehensive experiments have been carried out, reporting all possible combinations of source and target sets for each method. Note that the *Ideal-global* is not directly included in the table because its results can be inferred by selecting the best model for each target in the *Single-domain* experiments.

Concerning the baseline cases, i.e. *Single-domain* and *Multi-domain*, we observe that, as expected, the models trained with multiple sources are significantly more robust. Generally, we realize that the *Single-domain* method is not adequate for unsupervised experiments except in certain cases. For example, in the case in which SAL is the source, the supervised results achieve 96% of  $F_1$ , however, the results for the unsupervised ones are severely worsened, obtaining between 19.7% and 72.2% of performance. Note that this last figure matches with that obtained with EIN, which in turn, matches in document type with the source domain, since both contain music notation. Given that the remaining domains are collections of text manuscripts, important differences can be found in their content compared to the music documents. This might explain this high degradation in the binarization result. However, when a text manuscript—DIB or PHI—is used as the source, the degradation in the performance when dealing with music documents is not as accentuated. This may be attributed to the variability in the content of the images within the text collections, being the images of music documents more uniform. Therefore, this characteristic could be to blame for these disparate behaviors.

However, when multiple labeled domains are aggregated in the training process of a single domain (*Multi-domain*), the results are generally more robust and stable, with the exception of the two text manuscripts—DIB and PHI. When they are considered as unlabeled

**Table 2.** Average results in terms of  $F_1$  (%) for different source combinations. The “Avg. on target” column represents the average results among the target testing domains, i.e. the unsupervised experiments.

Method	Testing data				Avg. on target
	SAL	EIN	DIB	PHI	
<i>Single-domain</i>					
SAL	96.0	72.2	28.9	19.7	40.3
EIN	94.9	91.1	40.9	47.2	61.0
DIB	89.6	85.3	87.3	86.2	87.0
PHI	86.7	84.9	79.0	88.9	83.5
<i>Multi-domain</i>					
EIN-DIB-PHI	94.7	91.2	87.6	84.9	94.7
SAL-DIB-PHI	95.9	90.0	88.8	85.8	90.0
SAL-EIN-PHI	96.0	91.3	58.5	83.7	58.5
SAL-EIN-DIB	95.7	91.3	88.2	79.2	79.2
<i>Class-based selection</i>					
EIN-DIB-PHI	92.9	90.3	79.8	81.7	92.9
SAL-DIB-PHI	95.3	84.7	77.1	88.4	84.7
SAL-EIN-PHI	96.2	91.0	58.3	83.7	58.3
SAL-EIN-DIB	95.7	90.8	84.7	84.4	84.4
<i>Weighted combination</i>					
EIN-DIB-PHI	93.3	90.3	81.6	89.2	93.3
SAL-DIB-PHI	95.3	86.2	78.5	89.3	86.2
SAL-EIN-PHI	96.2	91.0	62.2	84.7	62.2
SAL-EIN-DIB	95.8	90.9	86.5	85.4	85.4
<i>Ideal-patch</i>					
EIN-DIB-PHI	95.2	91.5	89.0	91.0	95.2
SAL-DIB-PHI	96.4	88.7	88.7	90.9	88.7
SAL-EIN-PHI	96.6	91.5	80.8	89.1	80.8
SAL-EIN-DIB	96.6	91.5	88.4	86.6	86.6

beled domains in unsupervised experiments, they obtain the worst results found in *Multi-domain* ( $F_1 = 58.5\%$  and  $F_1 = 79.2\%$ , respectively). This situation can be attributed to the same reason explained in the *Single-domain*, i.e. because of the high variability of these text manuscripts.

With all this, we can conclude that the *Single-domain* is able to obtain good results for specific domains, but it is way less generalizable than *Multi-domain*.

With respect to our *Class-based selection* proposal, we observe also robust results, obtaining competitive results. Indeed, when one of the text manuscripts is unknown for the model—PHI—it achieves a  $F_1$  of 84.4% for the unique unlabeled domain—PHI—whereas the *Multi-domain* baseline provides a result of 79.2%, obtaining, therefore, a relative improvement of 6.6%. In the rest of the unsupervised experiments, the proposal obtains slight reductions in the performance between 0.3% and 5.8% of relative degradation, so that there is room for improvement, mainly associated with the domain classifier performance, which shall be later analyzed.

Concerning our second proposal—*Weighted combination*—the results indicate that this method is more promising. All the unsupervised cases obtain better performance with respect to *Class-based selection*, with relative improvements from 0.4% and 6.7% of  $F_1$ . In addition, compared with *Multi-domain*, both unsupervised experiments with a text manuscript as the target are significantly improved, with relative boosting of 6.3% and 7.8% for DIB and PHI, respectively. With respect to the music documents, the method also experiments certain degradation with relative figures between 1.5% and 4.2%. Therefore, with all this, the text manuscripts as the target domains may be consid-

ered as the main beneficiaries of our approach, whereas the music images do not find this boosting. Note, however, that the text manuscripts are those domains in which less performance is obtained by *Multi-domain*, so that, because of the high performance obtained in the music documents, this task becomes a major challenge to improve the baseline case.

Note that the cases in which our best proposal—*Weighted combination*—does not outperform the baselines are those where SAL and EIN (music corpora) are considered as testing data. Specifically, our proposal decreases the results for SAL from 94.7% to 93.3% and from 90% to 86.2% for EIN. These manuscripts are particularly difficult to process because of the high complexity of their content. They do not contain only one type of information, since there are staff lines, music notation, decorations and text. Precisely this variety is not provided by the text manuscripts considered in the experiments—DIB and PHI—so that the results of the individual models trained with these are not able to suitably detect such information and, therefore, the combination introduces errors. This leads us to conclude that the method will probably fail when few or none of the source domains are similar enough to the test one.

Furthermore, for the sake of the evaluation and analysis, the results that could be obtained if the domain classifier were always correct for each patch are also reported—*Ideal-patch*. We observe that, in this case, the unsupervised experiments are further improved, with relative figures for the text manuscript from 1.4% and 29.9% with respect to *Weighted combination*, representing a relevant improvement. If we compare the results with *Multi-domain*, we realize that the text domains experiment relative boosting between 9.3% and 38.1% in unsupervised learning, whereas concerning the music documents, the results reveal a relative improvement of 0.5% for SAL. Nevertheless, EIN do not exhibit the same situation, but a slight relative degradation of 1.4%.

**Table 3.** Average results in terms of  $F_1$  (%) on all targets in each approach considered. The bold figure highlights the best result in a real scenario, while the underlined figure indicates the best ideal case—which is only provided as a reference.

Method	$F_1$
<i>Baseline cases</i>	
Single-domain	68.0
Multi-domain	80.6
<i>Our approaches</i>	
Class-based selection	80.1
Weighted combination	<b>81.8</b>
<i>Ideal cases</i>	
Ideal-global	86.4
Ideal-patch	87.8

Although not all cases are improved by our approaches, or even by the *Ideal-patch* experiment, we observe that, according to the average result obtained on unknown target domains, our approaches, and especially *Weighted combination*, may be beneficial to the binarization task. The results shown in Table 3 support this idea, since the baselines *Single-domain* and *Multi-domain* achieve 68% and 80.6% of average performance in unsupervised experiments, and our best model *Weighted combination* obtains 81.8%, thus overcoming the baseline cases. In addition, the potential of our approach is proved with the ideal experiments, reporting that, if the domain classifier were perfect, the average

results could be improved to 86.4% for the case in which the domain classifier selects the best specialized model for the entire domain—*Ideal-global*—and 87.8% for *Ideal-patch*, i.e. the case in which the domain classifier selects the best model for each patch. The latter is especially relevant, obtaining a relative improvement of almost 9% with respect to the best baseline, and demonstrating that our approach could perform this task better as long as the way of selecting the specialized model is more accurate.

**Table 4.** Average results in terms of  $F_1$  (%) obtained by the domain classifier.

Sources	$F_1$
EIN-DIB-PHI	68.3
SAL-DIB-PHI	73.5
SAL-EIN-PHI	92.4
SAL-EIN-DIB	87.7

Since we saw that the domain classifier could be improved to closely match the results obtained in the *Ideal-patch* experiment, Table 4 shows a reference of the average performance provided by the domain classifier in controlled scenarios. We observe different behaviours depending on the domains involved. For example, the best case is presented when SAL, EIN and DIB are considered with a result of 92.4%. However, the worst result can be found when EIN, DIB and PHI are used, with only a performance of 68.3%. Indeed, the domain classifier reports the worst error rates when the two text manuscripts are considered, whereas the two cases in which the music domains SAL and EIN are used provide the best results. This reinforces the idea that the great variability of the text collections makes it difficult to carry out this classification task, likewise in the binarization performances, by adding an important degree of complexity in data. These values confirm that there exist great room for improvement that could be achieved with other neural architectures, more comprehensive hyper-parameter optimization or another combination mechanism.

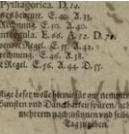
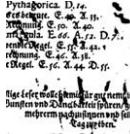
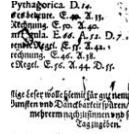
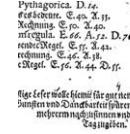
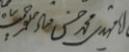
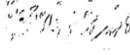
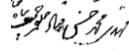
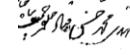
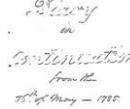
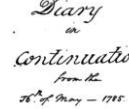
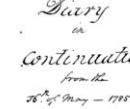
## 5. Qualitative Evaluation

To complement the previous experiments, we now report some representative examples of the benefits that our method can bring. Table 5 shows these selected examples, where we observe the potential of our approach compared with the *Multi-domain*, which is the most competitive baseline approach. Note that this table includes the binarization result by our best approach proposed in this work: *Weighted combination*.

Concerning the first example, the input image is a chunk from the DIB domain. After being processed by the *Multi-domain*, we observe that the method retrieves much of the foreground information, but also confuses several ink stains by labeling them as part of the foreground. These false-positive errors are slightly corrected by our method *Weighted combination*, whose result is a cleaner image. If we focus on the ideal case, we realize that the result that could be obtained by our approach is much closer to the ground-truth data, avoiding the aforementioned errors and obtaining a high-quality binarization result.

Concerning the two remaining examples, they correspond to images from PHI and DIB, respectively. Note the important difference of the input image in the third example compared with the first example, although both belong to the same domain. This explain

**Table 5.** Selected examples in which our approach is clearly beneficial for the binarization task. The columns, in order, represent the input images, the binarization performed by the best baseline method and our best approach, the ideal result that could be achieved and the ground truth, respectively.

Input	Multi-domain	Weighted combination	Ideal-patch	GT
1				
2				
3				

the great variability within DIB described in Section 4. We observe that the *Multi-domain* barely detect the ink strokes, missing a lot of ink pixels by categorizing them as background. These false negatives are severely corrected but our approach, obtaining a highly reliable result. In both cases, we realize that the differences between the ideal experiment and our approach are slightly perceptible, but at least visually, all the relevant information is detected. Indeed, only some details are missing if these results are compared with the ground truth.

Although the comparison of the performance in the last section does not seem very significant, these graphic examples prove that our approach can be beneficial for binarization. This reinforces our premise that specialized models may carry out the binarization more reliably than a generic model trained with multiple domains. Thus, our approach is able to combine the advantages of the specialized models with the generalization capability of the more generic ones, since our domain classifier contributes with this characteristic to our approach.

## 6. Conclusions

In this work, we present an alternative to improve the generalization of binarization strategies based on machine learning. Our methodology assumes a domain adaptation scenario, for which we have several training corpora (referred to as source domains) and our goal is to improve the performance of the binarization when applied to a domain for which no data is available (target domain). Our approach consists of training different neural networks for binarization, each of which is specifically trained with the data of a single source domain. We then provide a final binarization building upon the specialized models and the output of a domain classifier, which determines the suitability of each base model for the target sample.

In our experiments, we consider four datasets of different typology and it is shown that our starting hypothesis is quite promising. The upper bound that can be achieved by selecting a specialized model clearly improves the best baseline result, which consists of training a single model with all the data from the source domains. However, our way of

selecting the best model for each sample is not very competitive, since the improvement with respect to the aforementioned baseline is not negligible.

As future work, we plan to close the gap between our results and those achieved when assuming that the specialized model is correctly selected (*Ideal*). There are different ways of pursuing this goal. For example, instead of classifying a sample as to which domain it belongs to, we could train a model that predicts which of the specialized models binarize a sample better. This may lead to better results, given that the classifier learns how to choose a binarization model, instead of choosing a domain. In addition, other ways of selecting the specialized model could be considered by means of domain similarity and the domain2vec approach [22].

## Acknowledgements

This paper has been supported by Generalitat Valenciana through grant ACIF/2019/042 and project GV/2020/030, and Universidad de Alicante through project GRE19-04. The first two authors carried out this work as recipients of a grant from the Office for Educational Quality and Innovation of the University of Alicante, within the collaboration agreement with Banco de Santander S.A.

## References

- [1] He S, Wiering M, Schomaker L. Junction detection in handwritten documents and its application to writer identification. *Pattern Recognition*. 2015;48(12):4036-48.
- [2] Giotis AP, Sfikas G, Gatos B, Nikou C. A survey of document image word spotting techniques. *Pattern Recognition*. 2017;68:310-332.
- [3] Sulaiman A, Omar K, Nasrudin MF. Degraded Historical Document Binarization: A Review on Issues, Challenges, Techniques, and Future Directions. *Journal of Imaging*. 2019;5(4):48.
- [4] Otsu N. A threshold selection method from gray-level histograms. *Automatica*. 1975;11(285-296):23-7.
- [5] Niblack W. An introduction to digital image processing. Strandberg Publishing Company; 1985.
- [6] Sauvola J, Pietikäinen M. Adaptive document image binarization. *Pattern Recognition*. 2000;33(2):225-36.
- [7] Wolf C, Jolion JM, Chassaing F. Text localization, enhancement and binarization in multimedia documents. In: Proceedings of the International Conference on Pattern Recognition. vol. 2; 2002. p. 1037-40.
- [8] Gatos B, Pratikakis I, Perantonis SJ. Adaptive degraded document image binarization. *Pattern Recognition*. 2006;39(3):317-27.
- [9] Su B, Lu S, Tan CL. Robust document image binarization technique for degraded document images. *IEEE transactions on image processing*. 2013;22(4):1408-17.
- [10] Howe NR. A laplacian energy for document binarization. In: 2011 International Conference on Document Analysis and Recognition (ICDAR). IEEE; 2011. p. 6-10.
- [11] Kefali A, Sari T, Bahi H. Foreground-background separation by feed forward neural networks in old manuscripts. *Informatica*. 2014;38(4).
- [12] Calvo-Zaragoza J, Vigliensi G, Fujinaga I. Pixel-wise binarization of musical documents with convolutional neural networks. In: Fifteenth IAPR International Conference on Machine Vision Applications, MVA 2017, Nagoya, Japan, May 8-12, 2017; 2017. p. 362-5.
- [13] Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017;39(4):640-51.
- [14] Calvo-Zaragoza J, Gallego A. A selectional auto-encoder approach for document image binarization. *Pattern Recognition*. 2019;86:37-47.
- [15] Pratikakis I, Zagoris K, Karagiannis X, Tsochatzidis LT, Mondal T, Marthot-Santaniello I. ICDAR 2019 Competition on Document Image Binarization (DIBCO 2019). In: 2019 International Conference on

- Document Analysis and Recognition, ICDAR 2019, Sydney, Australia, September 20-25, 2019; 2019. p. 1547-56.
- [16] Castellanos FJ, Gallego AJ, Calvo-Zaragoza J. Unsupervised neural domain adaptation for document image binarization. *Pattern Recognition*. 2021;119:108099.
- [17] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics; 2010. p. 249-56.
- [18] Gatos B, Ntiogiannis K, Pratikakis I. ICDAR 2009 document image binarization contest (DIBCO 2009). In: 2009 10th International Conference on Document Analysis and Recognition. IEEE; 2009. p. 1375-82.
- [19] Ayatollahi SM, Ziae Nafchi H. Persian heritage image binarization competition (PHIBC 2012). In: 2013 First Iranian Conference on Pattern Recognition and Image Analysis (PRIA); 2013. p. 1-4.
- [20] Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. In: 3rd International Conference on Learning Representations. San Diego, USA; 2015. .
- [21] Bottou L. Large-scale machine learning with stochastic gradient descent. In: Proceedings of the International Conference on Computational Statistics. Springer; 2010. p. 177-86.
- [22] Peng X, Li Y, Saenko K. Domain2Vec: Domain Embedding for Unsupervised Domain Adaptation. In: Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part VI; 2020. p. 756-74.

# Control the Information of the Image with Anisotropic Diffusion and Isotropic Diffusion for the Image Classification

Hyun-Tae CHOI<sup>a</sup>, Nahyun LEE<sup>a</sup>, Jewon NO<sup>a</sup>, Sangil HAN<sup>a</sup>, Jaeho TAK<sup>a</sup>,  
Hwjin KIM<sup>a</sup>, Haegang LEE<sup>a</sup>, Seonghoon HAM<sup>a</sup> and Byung-Woo HONG<sup>a,1</sup>

<sup>a</sup>Computer Science Department, Chung-Ang University, Seoul, Korea

**Abstract.** Humans can recognize objects well even if they only show the shape of objects or an object is composed of several components. But, most of the classifiers in the deep learning framework are trained through original images without removing complex elements inside the object. And also, they do not remove things other than the object to be classified. So the classifiers are not as effective as the human classification of objects because they are trained with the original image which has many objects that the classifier does not want to classify. In this respect, we found out which pre-processing can improve the performance of the classifier the most by comparing the results of using data through other pre-processing. In this paper, we try to limit the amount of information in the object to a minimum. To restrict the information, we use anisotropic diffusion and isotropic diffusion, which are used for removing the noise in the images. By using the anisotropic diffusion and the isotropic diffusion for the pre-processing, only shapes of objects were passed to the classifier. With these diffusion processes, we can get similar classification accuracy compared to when using the original image, and we found out that although the original images are diffused too much, the classifier can classify the objects centered on discriminative parts of the objects.

**Keywords.** isotropic diffusion, anisotropic diffusion, classification, deep learning

## 1. Introduction

Even if the object is composed with many components or only the shape of the object is given, human can recognize what the object is. For the deep learning classification network, it is not easy to classify the object when only the object's shape is given, since most of the classifiers are trained with the original image. When we visualize the classification deep learning network with convolutional neural network, the network's layer which closer to an input layer, the more fined information is used and the last layer use more grained information [1]. Therefore, in order to accurately classify the object, the deep learning network uses both the shape and details of the object. In this respect, we want to find out how much object's information the network needed to classify the object. The simplest way to limit the amount of information is the diffusion. To diffuse the input

---

<sup>1</sup>Corresponding Author: Byung-Woo HONG, Computer Science Department, Chung-Ang University, Seoul, Korea; E-mail: hong@cau.ac.kr.

images, we use anisotropic diffusion [2] and isotropic diffusion. With this two diffusion process, we can limit the information of the objects by removing the artifacts inside of the objects. Also, with the diffused images, we can get similar classification accuracy compare to using the original images. In the remainder of this paper, we briefly present the two diffusing processes in Section 2 and then explain the our method for diffusion in Section 3. Next, we show our classification results in Section 4. Finally, we summarize our paper in Section 5.

## 2. Isotropic Diffusion and Anisotropic Diffusion

Isotropic diffusion and anisotropic diffusion algorithm reduce the noise in the image with a partial differential equation. The isotropic diffusion is the solution of the heat equation. When we set  $I(x,y,t)$  is the image at time  $t$  and the input image as  $I$ , the definition of the isotropic diffusion is Eq.(1).

$$\frac{\delta I(x,y,t)}{\delta t} = \Delta I \quad (1)$$

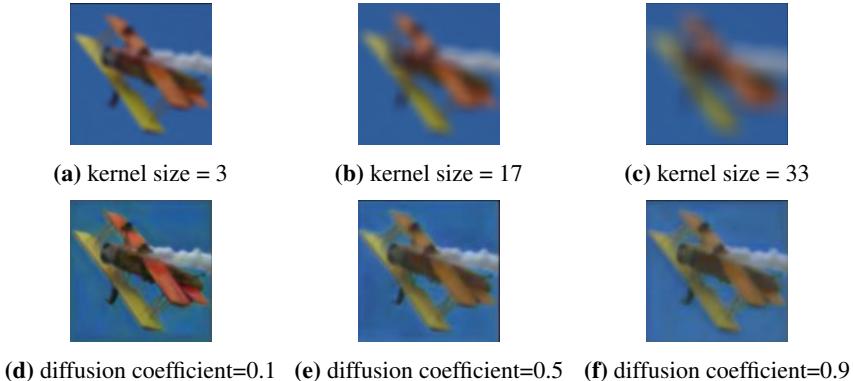
In Eq.(1),  $\Delta$  is the laplacian operator which is the inner product of  $\nabla$ ;  $\Delta = \nabla \cdot \nabla$  and  $\nabla$  is the gradient operator. But this isotropic diffusion algorithm diffuses the image uniformly in all directions, so when we apply the isotropic diffusion to the input image, it blurs too much that we cannot recognize the image. But unlike isotropic diffusioin, anisotropic diffusion algorithm diffuses the image with preserving the edge of the objects. The definition of the anisotropic diffusion is Eq.(2).

$$\frac{\delta I(x,y,t)}{\delta t} = \operatorname{div}(c(x,y,t)\nabla I) \quad (2)$$

Figure 1 shows the several diffused images. The kernel size of (a), (b) and (c) means the size of the gaussian kernel for the isotropic diffusion. When we apply large-sized kernel and big standard deviation, the gaussian kernel diffuses the image much more than the small-sized kernel. And the diffusion coefficient of (d), (e) and (f) is the  $\alpha$  of the Eq. 3. (a), (b) and (c) are the isotropic diffused image and (d), (e) and (f) are the anisotropic diffused image. (c) is the much more diffused than (a) and (f) is much mord diffused than (d). With (a), (b) and (c), we can know that when we diffuse the image with higher coefficient, we cannot recognize the image. However, with (d), (e) and (f), when we diffused the image with more higher coefficient, we can still recognize the object, because the anisotropic diffusion algorithm diffuses the image with preserving the edge of the objects.

## 3. Method

For the isotropic diffusion, we apply the gaussian kernel. When we use the small-sized and small standard deviation, the kernel diffuses the image little. On the contrary, with the big size kernel and big standard deviation, the kernel diffuses the image a lot. We use 8 kernel size; 3, 9, 15, 21, 27, 33, 39, 45. And set the standard deviation as twice of



**Figure 1.** (a), (b) and (c) show the isotropic diffused image and (d), (e) and (f) show the anisotropic diffused image.

each kernel size. We apply the isotropic diffusion before the classification network and the classification network is trained with the results of the isotropic diffusion.

Also, for the anisotropic diffusion, we make an auto-encoder with VGG-16 [3]. The encoder part is same as the VGG-16 except the fully connected layers and the decoder part is composed of the transposed convolution layers. And Eq.3 is used for the loss of the auto-encoder.

$$\begin{aligned} Loss &= \|U - I\|^2 + \alpha \|\nabla U\| \\ \|\nabla U\| &= \|\nabla U_x\| + \|\nabla U_y\| \end{aligned} \quad (3)$$

In Eq.3,  $I$  is the input and the  $U$  is the output of the anisotropic diffusion auto-encoder.  $\alpha$  is the diffusion coefficient and  $\nabla$  is the gradient. So,  $\nabla U_x$  is a differential  $U$  in the x-axis direction and  $\nabla U_y$  is a differential  $U$  in the y-axis direction. We apply the anisotropic diffusion as an auto-encoder network, before the classification network. So, the classifier is trained with  $U$  instead of the original input image  $I$ . For the diffusion coefficient  $\alpha$ , we use 7 diffusion coefficient; 0.001, 0.01, 0.1, 0.3, 0.5, 0.9. In Eq.3, the first term of the *Loss* is the data fidelity that control the difference between the output and the input image. And the second therm of the *Loss* is the total variation that diffuse the output image.

We use STL-10 [4] for the dataset and use VGG-16 [3], VGG-16 with batch normalization [5], VGG-19, VGG-19 with batch normalization, RESNET-18 [6] and RESNET-34. Also, we obeserve the class activation map [7] for each diffusion process and for each diffusion coefficient with the RESNET-34.

## 4. Results

To observe the two diffusion algorithms' influence to the information of the images, we check the classification score and class activation map. For the isotropic diffusion, we change the kernel size, and for the anisotropic diffusion, we change the  $\alpha$  in Eq.(3). The classification score is shown with Table 1. For the classification, we use 6 structures with

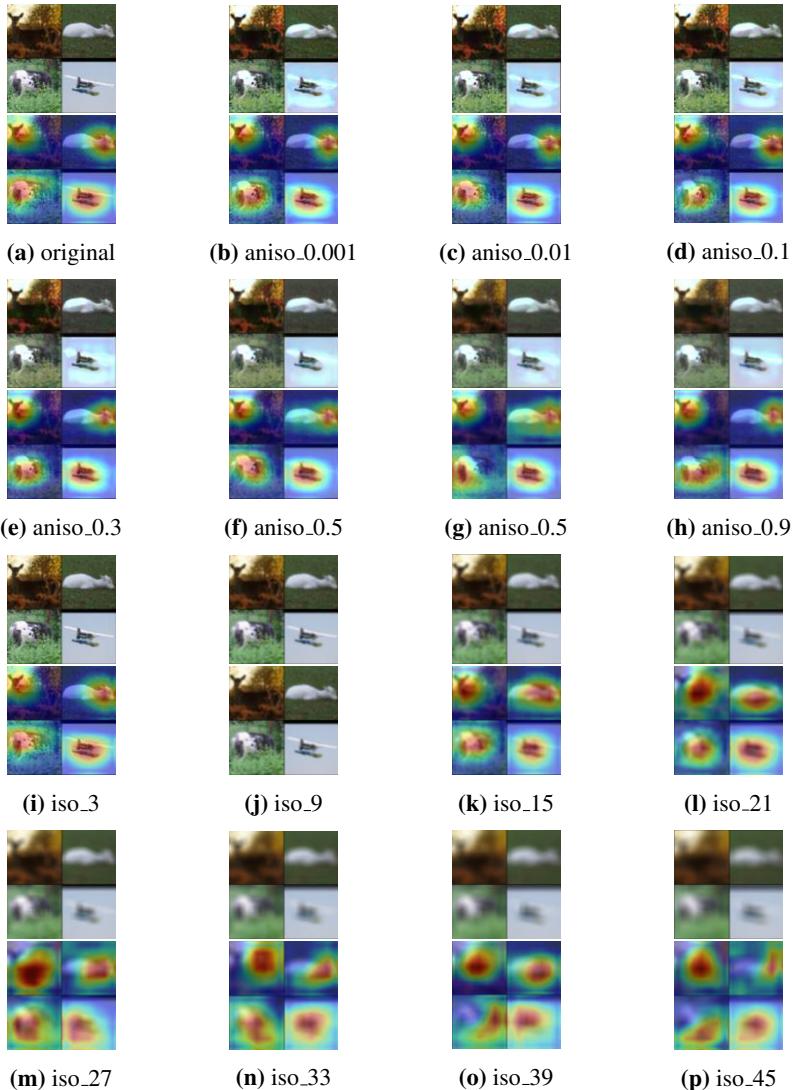
softmax; VGG-16, VGG-16.bn, VGG-19, VGG-19.bn, RESNET-18, RESNET-34. The VGG-16.bn means the VGG-16 with batch normalization and the VGG-19.bn means VGG-19 with batch normalization. We apply two diffusion algorithms; isotropic diffusion and anisotropic diffusion. For anisotropic diffusion, we use 8 kernel for each experiment. We represent the kernel size in the Table 1 as isotropic\_kernel size (e.g. isotropic\_3 means isotropic diffusion when the kernel size is 3.). And we represent the diffusion coefficient ( $\alpha$  in Eq.3) as aniso\_diffusion coefficient (e.g. aniso\_0.001 means anisotropic diffusion when the diffusion coefficient is 0.001.).

**Table 1.** classification accuracy(%) per each classifier, isotropic diffusion with 8 kernels size, and anisotropic diffusion with 7 diffusion coefficients. The **bn** means the batch normalization. We represent the kernel size of isotropic diffusion as **isotropic\_size** and the diffusion coefficient as **aniso\_coefficient**.

	VGG-16	VGG-16.bn	VGG-19	VGG-19.bn	RESNET-18	RESNET-34
<b>original</b>	<b>95.06</b>	<b>93.96</b>	<b>91.53</b>	<b>94.36</b>	<b>95.82</b>	<b>97.16</b>
isotropic_3	94.41	93.83	90.06	94.22	94.15	96.93
isotropic_9	89.25	86.98	83.35	88.07	89.51	93.4
isotropic_15	83.58	79.50	76.25	81.50	86.58	87.52
isotropic_21	79.21	74.08	73.07	74.56	81.06	82.86
isotropic_27	75.37	70.58	68.30	70.96	76.01	76.11
isotropic_33	71.15	67.12	68.53	67.93	68.80	70.86
isotropic_39	67.00	64.58	64.98	66.30	67.16	67.17
isotropic_45	61.56	63.05	62.23	63.56	61.42	64.82
aniso_0.001	92.33	91.51	89.08	92.57	93.57	95.05
aniso_0.01	92.50	91.85	88.57	92.53	93.53	94.88
aniso_0.1	91.42	91.30	88.07	92.15	93.22	94.58
aniso_0.3	90.28	89.06	86.96	90.53	92.28	93.57
aniso_0.5	89.31	88.56	85.00	89.55	92.05	92.93
aniso_0.5	89.05	87.56	83.45	88.21	90.85	92.18
aniso_0.9	87.57	86.53	82.13	87.22	90.30	91.15

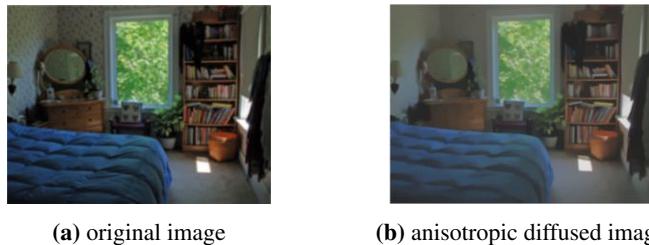
As we can see with Table 1, the isotropic diffusion algorithm has a negative effect on classification as diffusion degree becomes larger(i.e. the bigger kernel size). But, even if we apply the anisotropic diffusion with the properly large diffusion coefficient, we can get still high classification score. It means that the isotropic diffusion cannot preserve the information of the objects, but the anisotropic diffusion can preserve the information of the objects. This can also be seen through Figure 1. Also, Figure 2 show the class activation map of each diffusion algorithm and diffusion coefficient with RESNET-34.

In the Figure 2, each of the image sets, the top 4 images is the diffused images and the bottom 4 images is the class activation maps. **aniso\_k** represent the anisotropic diffusion with diffusion coefficient  $k$ (e.g. aniso\_0.001 means anisotropic diffusion when the diffusion coefficient is 0.001). **iso\_k** represent the isotropic diffusion with gaussian kernel size  $k$ (e.g. iso\_3 means isotropic diffusion when the gaussian kernel size is 3). When we compare the each results with the original image results (*a*), when we apply the anisotropic diffusion with various diffusion coefficient (*b*), (*c*), (*d*), (*e*), (*f*), (*g*), (*h*), the class activation maps do not change much. But, with the isotropic diffusion results (*i*), (*j*), (*k*), (*l*), (*m*), (*n*), (*o*), (*p*), we can see that the class activation maps get bigger as the kernel size gets bigger (i.e. much more diffusion). With, Table 1 and Figure 2, we



**Figure 2.** Diffused image and class activation map of each diffusion algorithms and diffusion coefficient. **aniso<sub>k</sub>** represent the anisotropic diffusion with diffusion coefficient  $k$ (e.g. aniso\_0.001 means anisotropic diffusion when the diffusion coefficient is 0.001). **iso<sub>k</sub>** represent the isotropic diffusion with gaussian kernel size  $k$ (e.g. iso\_3 means isotropic diffusion when the gaussian kernel size is 3).

can know that when we apply the properly large diffusion coefficient with anisotropic diffusion to the input image, the classifier can get similar classification accuracy even though the images are diffused. Also, with anisotropic diffusion we can decrease the noise of the image. Figure 3 shows the effect of anisotropic diffusion on the action of reducing noise of the image. In the Figure 3, the patterns on the wallpaper disappear.



(a) original image

(b) anisotropic diffused image

**Figure 3.** (a) is the original image and (b) is the anisotropic diffused image.

## 5. Conclusion

When the input image is diffused, there is a loss of information about the object. But if the object's shape is remained, the classification network still can classify the object. Two diffusion algorithm isotropic diffusion and anisotropic diffusion diffuse the image, but the isotropic diffusion cannot preserve the shape of the object well. On the other hand, anisotropic diffusion algorithm can preserve the edge that can preserve the shape of the object. With our experiment, the anisotropic diffusion algorithm is applied effectively to preserve the shape of the object. When we compare the anisotropic diffused images' classification score to original images, we can get similar classification score also, can get much more higher accuracy than using isotropic diffused images.

## Acknowledgment

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the National Program for Excellence in SW(20170001000051001) supervised by the IITP(Institute of Information & communications Technology PlanningEvaluation) in 2021

## References

- [1] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. European conference on computer vision. D. Fleet et al. (Eds.): ECCV 2014, Part I, LNCS 8689, 2014; pp. 818–833.
- [2] Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence, July 1990; 12(7): 629-639, doi: 10.1109/34.56205.
- [3] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv, 2014; abs/1409.1556.
- [4] Coates A, Ng A, Lee H. An analysis of single-layer networks in unsupervised feature learning. Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, in Proceedings of Machine Learning Research, 2011; 15:215-223 Available from <https://proceedings.mlr.press/v15/coates11a.html>.
- [5] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on Machine Learning, in Proceedings of Machine Learning Research, 2015; 37:448-456 Available from <https://proceedings.mlr.press/v37/ioffe15.html>.
- [6] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016; 770-778, doi: 10.1109/CVPR.2016.90.

- [7] Zhou BL, Khosla A, Lapedriza A, Oliva A and Torralba A. Learning deep features for discriminative localization. Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition. 2016; 2921-2929.

# Managing Study Stress of College Students Through Personality Traits

Ding DING <sup>a,†</sup>, Xinyue LIU <sup>b,†</sup> and Haoran XU <sup>c,1</sup>

<sup>a</sup> *North China Electric Power University, China*

<sup>b</sup> *School of Business, University of Wollongong, NSW 2522 Australia*

<sup>c</sup> *Alliance Business School, University of Manchester, Manchester, UK*

<sup>†</sup>*These authors contributed equally.*

**Abstract.** University students' stress management is a core topic in educational research; however, limited research has ever focused on how personality would impact students' perception of stress level. Thereby, this quantitative study, based on the Big Five Personality Trait Model, set out to investigates how openness, conscientiousness, extraversion, agreeableness, and neuroticism affect the college students' perception of their study stress. As a result of a survey with sixty college students in China, it is found that whilst students' demographic factors may not significantly predict their stress level, psychographic factors including conscientiousness, agreeableness and neuroticism can positively and significantly determine the college students' perception of stress. Based on the findings of the study, implications for future students' mental health management and ensuing educational research are put forward.

**Keywords.** Personality, Stress management, College students, Higher Education

## 1. Introduction

University life could be innately with high stress, manifested from paying for expensive tuition fees to taking exams and to looking for internships and jobs after graduation. Such overwhelming pressure, if handled inappropriately, could lead to negative consequences to students' physical and psychological conditions. In this situation, college students' stress management has been becoming an increasingly significant research domain for both academia and higher education practitioners [1].

Given the fact that the stress issue is prominent, there must be reasons for causing the stress. Previous studies have found that multifarious factors can influence the students' stress perception level, including but not limited to college course-related requirements, personal lifestyles and social connections [2]. Based on these factors, pertinent recommendations to release their stress have been simultaneously put forward, such as by noticing the symptoms, enhancing the social network, sleeping, eating well and doing more exercise [3].

Nonetheless, few extant studies have focused on how personal characteristics, or personality, impact students' stress levels. In accordance with previous psychology research, it can be said that personality traits could be more essential to explain the

---

<sup>1</sup> Corresponding author. Alliance Business School, University of Manchester, Manchester, UK; Email: haoran.xu-2@postgrad.manchester.ac.uk.

differences in stress level among college students [4]. In this context, this study attempts to investigate the mechanism between college students' personality and their stress levels. In the next section, based on previous literature, an array of hypotheses is made for further testing based on the collected empirical data.

## 2. Hypotheses Development

Personality can be defined as 'a set of distinctive traits and characteristics' and has been used as a valid predictor for human behaviors. In this study, personality would be selected as the independent variable to detect its influence on college students' stress levels.

To measure college students' personality traits, whilst there are distinct personality measurement models, the Big Five Personality Trait Model is one of the most implicative models frequently used to quantify ones' personality. Five dimensions included in the model are openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism [5].

Respectively, openness denotes how open a person is to new ideas and experiences; conscientiousness means how goal-directed, persistent, and organized a person is; extraversion is used to describe how much a person is energized by the outside world; Agreeableness relates to people displaying a capability of unselfishness, altruism and caring [6]; finally, neuroticism refers to one's tendency to face difficulty and stress with calmness, resolve, and security [7].

The Big Five model of personality is universally believed to be one of the most scientific and robust methods to measure personality differences. It is also the foundation of modern personality research [8] and is thereby adopted in this study to measure college students' personality traits. According to previous studies [9-10], these five factors can be hypothesized to be positively related to the factor of students' stress level, reacted as the dependent variable in this study. Finally, these hypotheses can be tested among college students in this study.

Aside from the five hypotheses with respect to personalities, two demographics factors are also included in the survey including sex and educational level, aiming to examine whether one's demography would impact the stress perception. Finally, the hypotheses made in this study are delineated in Table 1.

**Table 1.** Hypotheses made in this study.

Hypothesis	Description
H1	Gender would significantly impact college students' stress level
H2	Educational level would significantly impact college students' stress level
H3	Open to experience would significantly and negatively impact college students' stress level
H4	Conscientiousness would significantly and positively impact college students' stress level
H5	Extraversion would significantly and negatively impact college students' stress level
H6	Agreeableness would significantly and positively impact college students' stress level
H7	Neuroticism would significantly and positively impact college students' stress level

### 3. Methods

This study was conducted quantitatively, using a survey method to collect pertaining data to answer the research questions and hypotheses.

To begin with, with respect to the questionnaire design process, a questionnaire referring to the original big five personality was employed, resulting in a scale of one to five. These data would be used to measure their personality traits. As for the perceived stress level, a 5-point likert scale is used to measure students' perception of the study pressure. The questionnaire also included the data to collect students' demographics information including age, gender, major and educational level.

Regarding the data collection process, the survey questions were all transformed in the E-questionnaire using WeChat Form, which is an online survey platform and distributed through WeChat to Chinese higher education students. Then, these surveys were sent out using the convenience and snow-ball sampling methods, which is a type of non-probability method. As a result, 60 questionnaires collected without any missing data were included for data analysis.

As to the data analysis, the data were analysed through Excel and SPSS. More specifically, a descriptive analysis was conducted first to show the profile of the respondents. Then, mean tests, including independent sample t-test and one-way ANOVA were made to test the hypothesis one and two. After that, hypothesis three to seven are tested by the correlation test in SPSS. The results are shown in the next section.

### 4. Results

Table 2 shows the demographic information of the respondents. It is found that the collected data has covered the major demographic segments, showing that the sample collected can be used to implicate the population's (all the college students') conditions.

**Table 2.** Demographic profile of respondents.

Variable		Frequency (n=60)
Age	21	9
	22	10
	23	21
	24	15
	25	10
	26	5
Gender	Male	32
	Female	28
Educational level	Bachelor	30
	Master	22
	PhD	8

As for the mean test, independent sample t-test was conducted to detect whether gender can significantly impact college students' perceptions of stress (Table 3). The results show that gender had no significant impact on the college students' stress perception ( $p>0.05$ ).

**Table 3.** The mean test of student stress (gender)

	Levene's Test for Variances		T-test for Equality of Means							
	Equality		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence
Stress	Equal variances assumed	0.021	0.63		0.128	55	0.916	-0.0336	0.20234	-0.43456
	Not assumed				0.313	32.348	0.916	-0.0278	0.29067	-0.46677
										0.41098

Then, one-way Anova tests were performed to determine whether students with different educational levels would have distinct perceptions of study stress. The results reveal that there has been no significant difference among students with distinct educational background.

**Table 4.** One-way ANOVA (Education).

(I) EDUCATION	(J) EDUCATION	Mean	Difference	Std.	Sig.	95% Confidence Interval	
		(I-J)		Error		Lower Bound	Upper Bound
<b>Bachelor</b>	<b>Master</b>	0.45558		0.28802	0.119	-0.1205	1.0322
	<b>PhD</b>	0		0.30662	1	-0.6136	0.6136
<b>Master</b>	<b>Bachelor</b>	-0.45558		0.28802	0.119	-1.0322	0.1205
	<b>PhD</b>	-0.45558*		0.20655	0.031	-0.8692	-0.0426
<b>PhD</b>	<b>Bachelor</b>	0		0.30662	1	-0.6136	0.6136
	<b>Master</b>	.45558*		0.20655	0.031	0.0426	0.8692

\*The mean difference is significant at the 0.05 level.

Thus, both H1 and H2 needs to be rejected. In other words, the demographic factors are not effective predictors to forecast college students' study stress level. Eventually, correlation tests were made to test H3 to H7 (Table 5).

Based on the correlation test results, it is shown that conscientiousness, agreeableness, and neuroticism were statistically and significantly positively correlated with the perceived stress level, but openness and extraversion were not, which means that H3 and H5 should be rejected and H4, H6 and H7 can be initially accepted.

## 5. Discussion and Conclusion

In conclusion, as can be seen from the above results, this study firstly confirms that demographic factors are not significant predictors to monitor student's mental health conditions. In comparison, personalities are more valid and reliable to be considered when predicting college students' study stress levels. This finding is consistent with previous studies [11-12].

More specifically, this study also confirms that not all personality traits can be used to forecast students' study stress levels. As can be found in this article, three types of personality traits including conscientiousness, agreeableness and neuroticism would significantly impact one's stress level. Therefore, they should be adopted as more powerful indicators to predict college students stress levels.

**Table 5.** Correlation tests results.

		Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism	Stress
Openness	Pearson Correlation	1	0.567	0.555	0.689	0.567	.086
	Sig. (2-tailed)	0	0	0	0	0	0
	N	60	60	60	60	60	60
Conscientiousness	Pearson Correlation		1	0.655	0.98	0.777	.586**
	Sig. (2-tailed)		0	0	0	0	0
	N		60	60	60	60	60
Extraversion	Pearson Correlation			1	0.554	0.134	0.454
	Sig. (2-tailed)			0	0	0	0
	N			60	60	60	60
Agreeableness	Pearson Correlation				1	543	.70**
	Sig. (2-tailed)				0	0	0
	N				60	60	60
Neuroticism	Pearson Correlation					1	.720**
	Sig. (2-tailed)					0	0
	N					60	60
Stress	Pearson Correlation						1
	Sig. (2-tailed)						0
	N						60

\*\* Correlation is significant at the 0.01 level (2-tailed).

As a further illustration of the results, it can be implicated that when one student is more result-oriented and more sensitive toward others' opinions, he or she is more likely to experience mental health problems generated from study stress. In other words, if a student is more agreeable to the external environment, they tend to be more stressful. Other than that, students who show a more organized way of doing things and can persevere tend to feel more stress as well, as they have a stronger motivation to maintain or enhance a positive self-image and resist threats to it.

Moreover, among the Big Five personality, neuroticism has a negative effect on work pressure, leading to aggravation of emotional exhaustion. Students with high neuroticism tend to turn their faces with colleagues, which intensifies work pressure. Besides, conscientious people are hardworking, diligent, and self-disciplined, they tend to have an achievement striving motivation that is associated with a need to feel competent and accomplished [13]. Thus, students with high conscientiousness tend to have a higher level of ambition, focus more on their current main tasks, and work harder to achieve their goals, thus spontaneously avoiding irrelevant interference and pressure. However, since they can reconcile work and social, as the pressure from organizational factors increases, individuals continue to pay emotional resources, thus intensifying the pressure.

Finally, with respect to theoretical implication, it can be confirmed that the Big Five personality model is a representative of the theory of personality trait in personality psychology that can be used to probe stress perception. Hence, future theoretical construction should be more frequently based on ones' personality. Meanwhile, future

studies may harness other personality models to study what other personality characteristics are correlated with one's stress.

Concerning the practical implications, as personality plays key roles in determining students' study stress level, future college mental health consultation should pay more attention to how to construct students' positive personalities. For instance, the personality of openness can be further cultivated for college students, as it is worth noting that high-openness employees can often reduce their stress at work through social connections due to their outgoing personality, thereby relieving their mental tension in low-stress situations. Especially, in accordance with the results of the study, more mental guidance should be made toward those students who have a high standard for themselves.

Finally, there are some limitations in this study as well. First, this study only collected data from Chinese college students. Consequently, students from other cultural backgrounds are necessary to be included in future studies to come up with more generalized conclusions. More than that, even though the impact of personality toward the stress level is significant, there could exist more direct influential factors, or intermediaries, such as self-worth and self-efficacy. All in all, future studies can comprise more pertinent factors to dig into the college students stress management issues.

## References

- [1] Misra R, McKEAN M, West S, Russo T. Academic stress of college students: comparison of student and faculty perceptions. *College Student Journal*. 2000;34(2): 236-245.
- [2] Villanova P, Bownas DA. Dimensions of College Student Stress J. 1984.
- [3] Feld L D, Shusterman A. Into the pressure cooker: Student stress in college preparatory high schools. *J Adolescence*. 2015; 41: 31-42.
- [4] Judge T A, Higgins C A, Thoresen C J, Barrick M Ret. The big five personality traits, general mental ability, and career success across the life span. *J Personnel psychology*. 1999;52(3): 621-652.
- [5] Gosling S D, Rentfrow P J, Swann Jr W B. A very brief measure of the Big-Five personality domains. *J. Res. Personality*. 2003;37(6): 504-528.
- [6] Syed A R, Rehman K U, Kitchlew N. Impact of perceived leadership style on employees' work stress: Moderating and mediating role big 5 personality traits. *J Paradigms*. 2018;12(1): 6-15.
- [7] Jansi A M, Anbazhagan S. The relationship between big 5 personality traits and life satisfaction of among ncc women students. *International J. Management*. 2017;8(2).
- [8] Saklofske D H, Austin E J, Mastoras S M, Beaton L, Osborne S E. Relationships of personality, affect, emotional intelligence and coping with student stress and academic success: Different patterns of association for stress and success. *J Learning Individual Differences*. 2012;22(2): 251-257.
- [9] Tosevski D L, Milovancevic M P, Gajic S D. Personality and psychopathology of university students. *J Current Opinion in Psychiatry*. 2010;23(1): 48-52.
- [10] Bagherian S A, Mohammadian F, Azadi R. The relation between personality type and stress. *AETAS – J. History Related Disciplines*. 2015;(4): 21.
- [11] Han W L, Kim J E. A study on university students' personality traits and resilience. *J. Emotional Behavioral Disorders*. 2017. 33(3): 21-40.
- [12] Kato T. Styles of handling interpersonal conflict, personality, and mental health in undergraduate students. *Japanese J. Social Psychology*. 2013;18(2): 78-88
- [13] Holman D J, Hughes D J. Transactions between Big - 5 personality traits and job characteristics across 20 years. *J Occupational and Organizational Psychology*. 2021.

# An Adaptive Network Model of the Role of the Gut-Brain Axis in Parkinson's Disease

Catherine GOETZINGER<sup>a, b</sup>, Korinzia TONILO<sup>c, 1</sup> and Jan TREUR<sup>d</sup>

<sup>a</sup> Deep Digital Phenotyping Research Unit, Department of Population Health,  
Luxembourg Institute of Health, Strassen, Luxembourg

<sup>b</sup> University of Luxembourg, Faculty of Science, Technology and Medicine,  
Luxembourg

<sup>c</sup> Department of Management, University of Bologna, Italy

<sup>d</sup> Social AI Group, Vrije Universiteit Amsterdam

**Abstract.** This paper presents an adaptive network model of the brain-gut axis and its related mechanisms that play major roles in the development of Parkinson's Disease. Simulations gave useful insight into how the biological and mental pathways interact with each other. In addition, the model provides information on the different time spans that are taken into consideration within these processes.

**Keywords.** Brain-Gut axis, Parkinson's Disease, adaptive network model, self-modeling network.

## 1. Introduction

Parkinson's Disease (PD) is a systemic disease clinically defined by the degeneration of dopaminergic neurons in the brain. It has become the second most widespread neurodegenerative disorder in the world, expected to affect more than 10 million people worldwide by 2030 [1, 2].  $\alpha$ -synucleinopathy and motor impairment are two of the main signals of PD pathogenesis. They occur when 55% of the dopaminergic neurons are damaged in the substantia nigra, the area of the midbrain containing the nerve cells producing dopamine. Besides motor impairment, over time PD patients can show metabolic imbalance, with half of them experiencing constipation before the outset of other clinical features [3]. Recently, some studies have begun to explore the mechanisms that connect the gut and the brain. For example, changes in the composition of intestinal microbiota have been associated with neurological and neurodevelopmental disorders [4].

In this paper, we will specifically focus on the influence of gut microbiota mechanisms on PD mechanisms, analyzing computationally how such processes take place. This computational analysis addressing the different processes at different time scales involved, was performed based on multi-order adaptive networks modeled by the concept of self-modeling network.

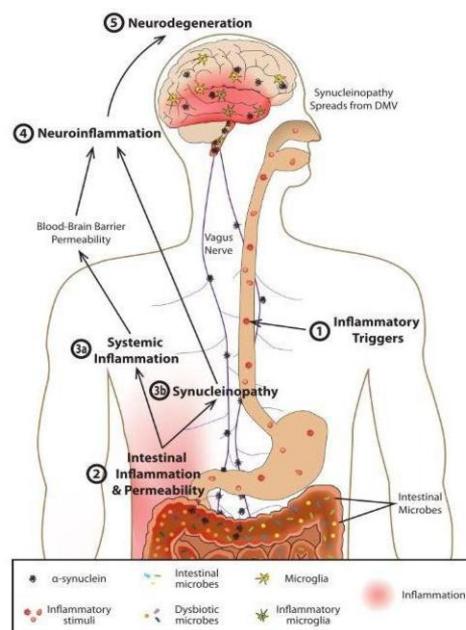
---

<sup>1</sup> Corresponding Author: Korinzia TONILO, Department of Management, University of Bologna, Italy; E-mail: [korinzia.toniolo2@unibo.it](mailto:korinzia.toniolo2@unibo.it).

## 2. Background Literature

The microbiota is composed by trillions of microorganisms living within our bodies. In the past decade, researchers have started to study it as one of the key regulators of the gut-brain axis, recognizing its important role for the central nervous system (CNS) [5]. The gut-brain axis is a bidirectional interaction, which integrates the cognitive and emotional centers of the brain with peripheral gastrointestinal mechanisms [3]. Recent studies on the gut-brain axis have revealed multiple ways the gut microbiota composition can affect people's daily lives and disease outcomes. Multiple factors can affect the microbiota composition over time, including infections, antibiotic medications, environmental stressors, and host genetics [6]. For example, it has been found that the microbiota of a PD patient tends to be characterized by reduced carbohydrate fermentation and butyrate synthesis capacity, increased proteolytic fermentation and production of deleterious amino acid metabolites, including p-cresol and phenylacetylglutamine [7]. Further, PD patients' microbiota shows reduced levels of fecal short-chain fatty acids that may generate alteration in the enteric nervous system, causing constipation in PD patients [8]. Evidence from laboratories' investigation support the hypothesis of the connection between the complexity of gut microorganisms and PD pathogenesis [9–11].

Figure 1 provides an overview of how the connection between gut and brain mechanisms may occur in PD pathogenesis.



**Figure 1.** Model of gut-originating, inflammation-driven PD pathogenesis adopted from (4).

An inflammatory trigger, for example, a toxin or the intake of antibiotics, can initiate an inflammatory response in the gut and contribute to developing bad gut microbiota. The gut microbial toxins may then trigger the production of an increased aggregation of  $\alpha$ -synuclein ( $\alpha$ -syn) protein in the enteric nervous system (ENS), which can be transmitted to the central nervous system (CNS) through the vagus nerve. The presence

of high aggregations of the  $\alpha$ -synuclein protein (synucleinopathy) in the gut is the main manifestation of PD. Peripheral inflammation can then trigger a systemic inflammatory response, increasing the likelihood of neuroinflammation and accelerating the pace of neurodegeneration.

In this context, some studies have started to investigate the use of probiotics for Parkinson's disease prevention or treatment [12]. The presence of probiotic bacteria favors a healthy gut environment by modulating the host immune response and balancing the intestinal microflora.

Nutraceutical interventions aim to normalize the gut microbiome dysbiosis and improve biological outcomes in PD conditions. In this sense, regular intakes of probiotics can positively affect the gut microflora by increasing the presence of beneficial microorganisms [13]. Thus, the use of probiotics together with healthy diet habits can be seen as an adjuvant therapy to the use of standard PD medications, such as Levodopa. Brain cells in the substantia nigra within PD patients no longer produce the chemical dopamine, essential for brain neurons to communicate and control movement [14]. The chemical imbalance in PD causes physical symptoms such as resting tremors, stiffness, and slowness of movements, among others. Levodopa, a central nervous system agent, helps people with PD by replacing the chemical messenger dopamine [15]. It helps to alleviate the symptoms of PD by providing a supply of dopamine. It has to be noted, however, that Levodopa does not slow or reduce the progression of PD. Levodopa is an amino acid that is processed to dopamine in the brain. Today, Levodopa is the most effective drug used by PD patients [14].

### 3. Network-Oriented Modeling

To describe and analyze the influence of gut microbiota mechanisms on PD mechanisms computationally, we used a network-oriented modeling approach as described in [16, 17].

The introduced network model will model both the dynamics of the interaction processes within networks and of network evolution. Within PD we have processes that have very different time spans; some of the processes do occur on a daily basis (e.g., intake of fiber levels), whereas others only occur over the years (e.g., neurodegeneration). This type of model is declarative and characterized by connectivity characteristics, aggregation characteristics, and timing characteristics [17]:

- **Connectivity characteristics:** connections from states  $X$  to  $Y$ , having connection weights  $\omega_{X,Y}$  specifying their strengths
- **Aggregation characteristics:** each state  $Y$  has a combination function  $c_Y$  that specifies how impact from all incoming connections on  $Y$  is aggregated. Based on a list of basic combination functions  $bcf_i$  (each with some parameters) provided by an available library, such a combination function can be specified by weights  $\gamma_i$  and parameters  $\pi_{ij}$  for these basic combination functions  $bcf_i$
- **Timing characteristics:** each state  $Y$  has a speed factor  $\eta_Y$  specifying how fast  $Y$  changes

Furthermore, these models can be conceptualized in the form of matrices or labeled graphs. These conceptual graphs illustrate states and their connections between them plus some labels. States have varying activation levels over time. Moreover, the concept of weights of a connection is used for the strengths of impact of one state on another one.

Finally, combination functions are used that represent the aggregation of multiple causal impacts on a state, and speed factors to represent the speed of change of a state, thus modeling the timing of the process. Next to matrices and labeled graphs, models can be formalized in a numeric format:

$$Y(t+\Delta t) = Y(t) + \eta_Y [\mathbf{c}_Y(\omega_{X_1,Y} X_1(t), \dots, \omega_{X_k,Y} X_k(t)) - Y(t)] \Delta t \quad (1)$$

Eq. (1) represents how the states  $X_1$  to  $X_k$  from which  $Y$  gets incoming connections affects the activation level of  $Y$ . The effect of the combination function  $\mathbf{c}_Y(\cdot)$  on  $Y$  is exerted gradually over time, depending on speed factor  $\eta_Y$ . Eq. (1) are hidden in the software environment that can be used for simulation and analysis; see [17]. Within the above-mentioned software, more than 45 basic combination functions are included. The combination functions used in the present paper are presented in Table 1.

**Table 1.** The basic combination functions from the library used in the presented model.

	Notation	Formula	Parameters
Stepmod	<b>stepmod<sub>p,δ</sub>(V)</b>	0 if $t \bmod p \leq \delta$ , else 1	$p$ repetition interval length, $\delta$ step time
Advanced logistic sum	<b>alogistic<sub>σ,τ</sub>(V<sub>1</sub>, ..., V<sub>k</sub>)</b>	$\frac{1}{1+e^{-\sigma(V_1+\dots+V_k-\tau)}} - \frac{1}{1+e^{\sigma\tau}}](1+e^{-\sigma\tau})$	Steepness $\sigma > 0$ Excitability threshold $\tau$

#### 4. Self-Modeling Networks for Adaptivity

‘Network characteristics’ and ‘network states’ are by definition two distinct concepts for a network. Self-modeling networks are based on a self-model concept for part of their own network structure. A self-model represents certain network structure characteristics (e.g., connection weights or excitability thresholds) by network states. Hence self-models extend the states from the base level by additional states at a higher level, enabling adaptation of the network structure. This step is also called network reification. It is possible to create higher-order self-modeling levels, where network characteristics from one level are related to states from another level: second-order or higher-order adaptive networks [17].

In our scenario, self-modeling is added for those network characteristics that experience a change over time. Therefore, we end up with different levels of the model that interact with each other; a network model for the base network and its within-network dynamics and a numerical model for the adaptivity of (some of) the network structure characteristics of the base network [17]. Self-modeling uses a network-oriented conceptualization similar to what was described above.

- **Connectivity self-model**

Self-model states  $\mathbf{W}_{X,Y}$  are used to represent connectivity characteristics, i.e., connection weights  $\omega_{X,Y}$

- **Aggregation self-model**

Self-model states  $\mathbf{C}_{j,Y}$  are used to represent the first type of aggregation characteristics: combination function weights  $\gamma_{j,Y}$ . Self-model states  $\mathbf{P}_{i,j,Y}$  are used to represent the second type of aggregation characteristics: combination function parameters  $\pi_{i,j,Y}$ .

- **Timing self-model**

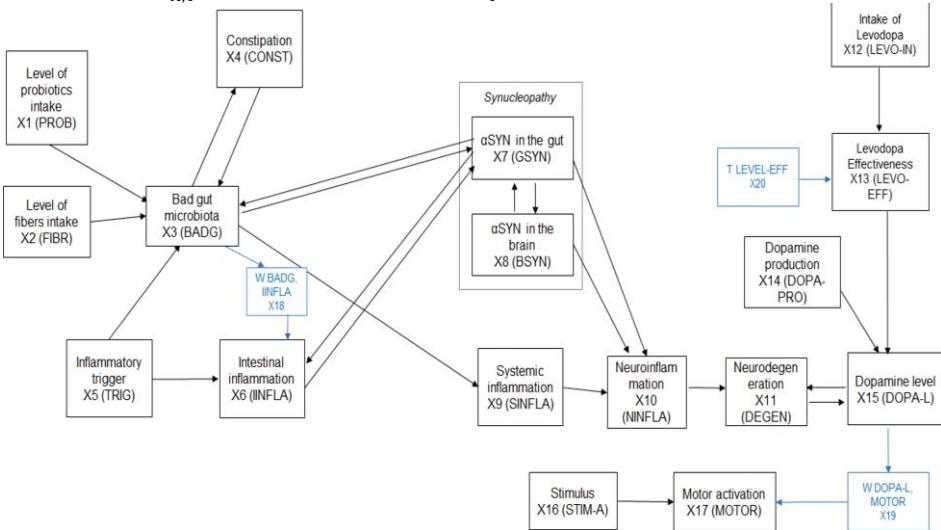
Self-model states  $\mathbf{H}_Y$  are used to represent timing characteristics: speed factors  $\eta_Y$

In certain cases, the names using the letters **W**, **C**, **P** and **H** can be chosen in a different, more specific manner for example, **T** for excitability threshold parameter  $\tau$ .

The self-modeling construction can be iterated so that second-order (or even higher-order) adaptation can be modeled. An example of a second-order self-model state is a state  $\mathbf{Hw}_{X,Y}$  that represents the speed factor (adaptation rate)  $\eta$  of (first-order) self-model state  $\mathbf{W}_{X,Y}$ . This type of self-model states can be used in an adaptive network model to control the adaptation: decreasing activation values of  $\mathbf{Hw}_{X,Y}$  model make the adaptation slow down, and a value 0 causes a complete freezing of the adaptation.

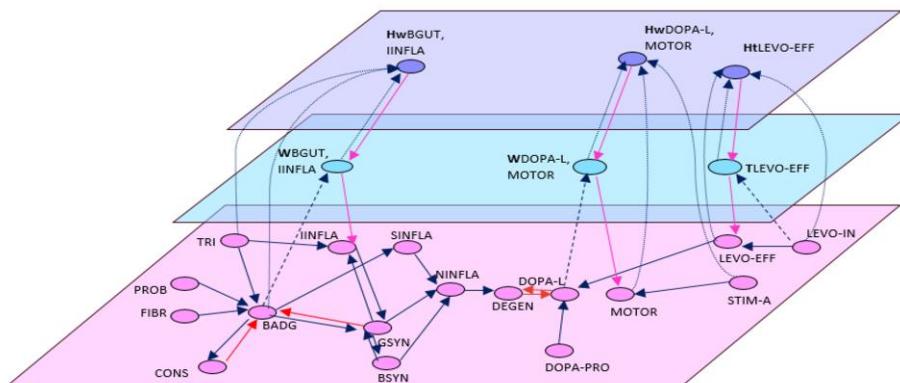
## 5. An integrative Adaptive Network Model

Moving from the general premises, we first developed the conceptual model presented in Figure 2, which shows the connections between the gut mechanisms (on the left side of the graph) and the brain mechanisms (on the right side of the graph) which processes play a major role in PD pathogenesis. The blue boxes in Figure 2 represent two self-model states  $\mathbf{W}_{X,Y}$  and one self-model state  $\mathbf{T}_Y$ .



**Figure 2 - Conceptual Model for the brain-gut axis in PD.**

The designed self-modeling network model integrates both biological (physiological) and mental (neural) pathways. Both pathways interact in the so-called brain-gut axis which is represented at the base level (the pink plane in Figure 3). In PD, processes occur at different time spans and take a different amount of time to impact other processes. Therefore, an important addition to this model contributed here is an adaptive form of timing. Figure 3 represents a graphical 3D format that shows the base model on the bottom, and the self-model states are depicted at the next levels (reification levels).



**Figure 3.** 3D figure of the connectivity of an integrative adaptive network model for the gut-brain axis in PD.

The states are labeled and described in Table 2. On the base level (pink plane) ‘PROB’, ‘FIBR’, ‘TRI’, ‘DOPA-PRO’ and ‘LEVO-IN’ are independent states that have no incoming influences yet influence other states. These independent states are time-dependent and occur occasionally; therefore, the stepmod combination function was used for these states to generate repetitive patterns. For all the other states, we used the allogistic combination function.

On the base level, some arrows are red, meaning that they have a negative influence on the respective state. Through literature review, we found that constipation and  $\alpha$ -syn in the gut have a negative influence on the gut microbiota. Furthermore, we noted that neurodegeneration and dopamine levels negatively influence each other. Meaning that low dopamine levels negatively affect neurodegeneration, and with increased neurodegeneration, it has a negative impact on dopamine levels.

Reading this 3D figure, we can see that probiotic ( $X_1$ ) and fiber ( $X_2$ ) intake influence the gut microbiota. If you have low levels of both, you tend to develop a bad gut microbiota ( $X_3$ ). In addition, inflammatory triggers ( $X_5$ ) such as toxins influence the bad gut composition. Having a bad gut composition creates constipation ( $X_4$ ), which in return negatively affects the bad gut composition. Moreover, the bad gut composition has an influence on the system inflammation ( $X_9$ ) as well as on the  $\alpha$ -syn in the gut ( $X_7$ ).

The synucleinopathy ( $X_7$  and  $X_8$ ) and the intestinal inflammation ( $X_6$ ) affect each other. In addition, synucleinopathy influences neuro-inflammation ( $X_{10}$ ), which means that neurodegeneration ( $X_{11}$ ) is triggered. In addition, neurodegeneration increases by low levels of dopamine chemicals ( $X_{15}$ ). The latter is affected by the production of dopamine ( $X_{14}$ ) in the brain.

As stated in the background section, Levodopa is the most common drug used in PD. It helps to increase the level of dopamine in the brain. Therefore, levodopa intake ( $X_{12}$ ) and its efficacy ( $X_{13}$ ) affect dopamine levels. On an everyday basis, individuals receive stimulus for activation ( $X_{16}$ ), which by consequences, therefore, triggers your motor activation ( $X_{17}$ ).

As already mentioned in the previous section, the present models depicting the gut-brain axis in PD deals with states occurring at various time scales. Intake of inflammatory triggers can happen on a regular occasion, whereas the bad gut microbiota composition happens throughout a longer period. In order to adjust for such dynamics, the model includes time regulations. In the 3D figure, on the first adaptive level (blue plan), a weight connection was added to affect the inflammation trigger connection on intestinal

inflammation ( $X_6$ ). This is done by adding state  $\mathbf{W}_{\text{BADG}, \text{IINFLA}}$  (1<sup>st</sup> order self-model level) into our model. Bad gut composition makes you more sensible for intestinal inflammation and thus strengthens the link between the inflammatory trigger and intestinal inflammation. Therefore,  $\mathbf{W}_{\text{BADG}, \text{IINFLA}}$  ( $X_{18}$ ) will regulate the impact on how much inflammatory triggers influence intestinal inflammation. Another weight connection on this first adaptation level was created between state  $X_{12}$  and  $X_{17}$ ,  $\mathbf{W}_{\text{DOPA-L}, \text{MOTOR}}$  ( $X_{19}$ ). The dopamine level makes you more sensible for motor activation. The added  $\mathbf{W}_{\text{DOPA-L}, \text{MOTOR}}$  state, therefore, strengthens the link between the stimulus for activation state and motor activation state. Over time, the effectiveness of Levodopa decreases even, if the intake of the drug increases. In order to adjust for this dynamic process in the current model, an adaptive excitability threshold  $\mathbf{T}_{\text{LEVEL-EFF}}$  ( $X_{20}$ ) was added to weaken the effectiveness of Levodopa over time.

On the second level of the model, the **H**-states were added as a form of speed control for adaptation. Therefore  $\mathbf{Hw}_{\text{BADG}, \text{IINFLA}}$  ( $X_{21}$ ) will regulate the learning speed of  $\mathbf{W}_{\text{BADG}, \text{IINFLA}}$ . Similarly,  $\mathbf{Hw}_{\text{DOPA-L}, \text{MOTOR}}$  ( $X_{22}$ ) will regulate the learning speed of  $\mathbf{W}_{\text{DOPA-L}, \text{MOTOR}}$  and  $\mathbf{Ht}_{\text{LEVEL-EFF}}$  ( $X_{23}$ ) will regulate the learning speed of  $\mathbf{T}_{\text{LEVEL-EFF}}$ , respectively.

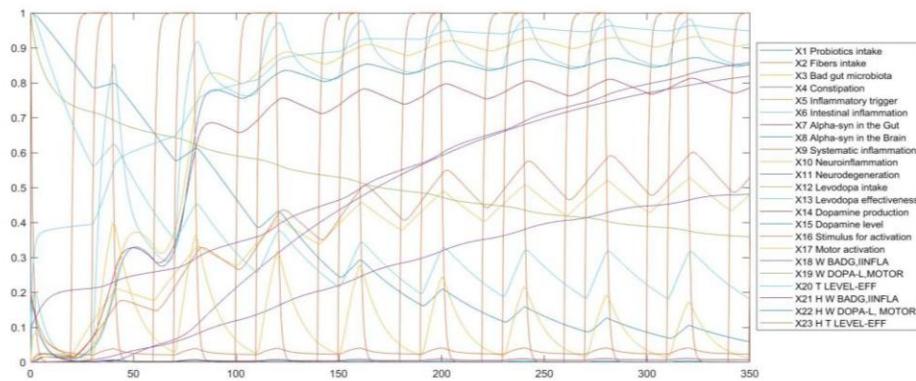
**Table 2.** State labels and description

States	Label	Description
$X_1$	PROB	Probiotics intake
$X_2$	FIBR	Fibers intake
$X_3$	BADG	Bad gut microbiota
$X_4$	CONST	Constipation
$X_5$	TRIG	Inflammatory trigger
$X_6$	IINFLA	Intestinal inflammation
$X_7$	GSYN	$\alpha$ -syn protein in the Gut
$X_8$	BSYN	$\alpha$ -syn protein in the Brain
$X_9$	SINFLA	Systemic inflammation
$X_{10}$	NINFLA	Neuroinflammation
$X_{11}$	DEGEN	Neurodegeneration
$X_{12}$	LEVO-IN	Levodopa intake
$X_{13}$	LEVO-EFF	Levodopa effectiveness
$X_{14}$	DOPA-PRO	Dopamine production
$X_{15}$	DOPA-L	Dopamine level
$X_{16}$	STIM-A	Stimulus for activation
$X_{17}$	MOTOR	Motor activation
$X_{18}$	$\mathbf{W}_{\text{BADG}, \text{IINFLA}}$	Self-model state for the weight of the connection from BADG to IINFL
$X_{19}$	$\mathbf{W}_{\text{DOPA-L}, \text{MOTOR}}$	Self-model state for the weight of the connection from DOPA-L to MOTOR
$X_{20}$	$\mathbf{T}_{\text{LEVEL-EFF}}$	Self-model state for the excitability threshold of LEVEL-EFF
$X_{21}$	$\mathbf{Hw}_{\text{BADG}, \text{IINFLA}}$	Self-model state for adaptation speed of $\mathbf{W}_{\text{BADG}, \text{IINFLA}}$
$X_{22}$	$\mathbf{Hw}_{\text{DOPA-L}, \text{MOTOR}}$	Self-model state for adaptation speed of $\mathbf{W}_{\text{DOPA-L}, \text{MOTOR}}$
$X_{23}$	$\mathbf{Ht}_{\text{LEVEL-EFF}}$	Self-model state for adaptation speed of $\mathbf{T}_{\text{LEVEL-EFF}}$

## 6. Simulation of the Integrative Adaptive Network Model

Within the present section, we will discuss the findings of the simulation that was generated using the software environment, and the specification by role matrices, detailing the network characteristics, which are presented in the Appendix available as Linked Data at URL <https://www.researchgate.net/publication/353764251>. Figure 4 shows the simulation representing the brain-gut axis process in PD. This scenario shows

a adaptation effect after the inflammatory trigger intake, the stimulus activation, and the Levodopa effectiveness.



**Figure 4.** Simulation of the adaptive network model.

The simulation shows that probiotics and fiber levels continuously decrease over time due to an increasingly bad gut composition. With this, constipation slowly worsens. States  $X_{21}$ ,  $X_{22}$ ,  $X_{23}$ , the adaptive connection weights and excitability threshold are very slow in the present model as can be spotted at the bottom of the simulation. Due to a slow adaptation rate of the model ( $\mathbf{W}_{\text{BADG}, \text{INFLA}}$ ), intestinal inflammation only expresses after a certain amount of time once the influence of the inflammatory trigger is high enough to affect the intestinal inflammation. The same holds for the Motor activation that slowly decreases over the time span ( $\mathbf{W}_{\text{DOPA-L}, \text{MOTOR}}$ ). The excitability threshold ( $T_{\text{LEVEL-EFF}}$ ) is slow at the beginning of the simulation and increases after 100-time points. Therefore, the Levodopa effectiveness remains high at the beginning of the simulation and thus decreases with increasing  $T_{\text{LEVEL-EFF}}$ .

## 7. Discussion

The presented model and its respective simulation is the first computational representation of the brain-gut axis processes and mechanisms that play major roles in PD. This simulation is useful because it gives insight into how the biological and mental pathways interact with each other. In addition, the model provides information on the different time spans that are taken into consideration within these PD processes.

Nevertheless, the present model remains a simplified version of the PD pathogenesis and further work in computational modeling could consider, for example, the impact of the mitochondria process in PD. Furthermore, future studies may have a closer look at specific probiotics and their influence on the brain-gut axis in PD. These could eventually be helpful in the prevention strategies for PD.

## References

- [1] Burgueño JF, Abreu MT. Epithelial Toll-like receptors and their role in gut homeostasis and disease. *Nat Rev Gastroenterol Hepatol.* 2020 May;17(5):263–78.
- [2] Marras C, Beck JC, Bower JH, Roberts E, Ritz B, Ross GW, et al. Prevalence of Parkinson's disease across North America. *npj Parkinson's Disease.* 2018 Jul 10;4(1):1–7.
- [3] Dogra N, Mani RJ, Katare DP. The gut-brain axis: two ways signaling in Parkinson's disease. *Cell Mol Neurobiol [Internet].* 2021 Mar 1 [cited 2021 Sep 14]; Available from: <https://doi.org/10.1007/s10571-021-01066-7>
- [4] Houser MC, Tansey MG. The gut-brain axis: is intestinal inflammation a silent driver of Parkinson's disease pathogenesis? *npj Parkinson's Disease.* 2017 Dec;3(1):3.
- [5] Wang Y, Kasper LH. The role of microbiome in central nervous system disorders. *Brain Behav Immun.* 2014 May; 38:1–12.
- [6] Cryan JF, O'Riordan KJ, Cowan CSM, Sandhu KV, Bastiaanssen TFS, Boehme M, et al. The microbiota-gut-brain axis. *Physiological Reviews.* 2019 Oct 1;99(4):1877–2013.
- [7] Cirstea MS, Yu AC, Golz E, Sundvick K, Kliger D, Radisavljevic N, et al. Microbiota composition and metabolism are associated with gut function in Parkinson's disease. *Mov Disord.* 2020 Jul;35(7):1208–17.
- [8] Koh A, De Vadder F, Kovatcheva-Datchary P, Bäckhed F. From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites. *Cell.* 2016 Jun 2;165(6):1332–45.
- [9] Bedarf JR, Hildebrand F, Coelho LP, Sunagawa S, Bahram M, Goeser F, et al. Functional implications of microbial and viral gut metagenome changes in early stage L-DOPA-naïve Parkinson's disease patients. *Genome Medicine.* 2017 Apr 28;9(1):39.
- [10] Hill-Burns EM, Debelius JW, Morton JT, Wissemann WT, Lewis MR, Wallen ZD, et al. Parkinson's disease and Parkinson's disease medications have distinct signatures of the gut microbiome. *Mov Disord.* 2017 May;32(5):739–49.
- [11] Keshavarzian A, Green SJ, Engen PA, Voigt RM, Naqib A, Forsyth CB, et al. Colonic bacterial composition in Parkinson's disease. *Mov Disord.* 2015 Sep;30(10):1351–60.
- [12] Castelli V, d'Angelo M, Quintiliani M, Benedetti E, Cifone MG, Cimini A. The emerging role of probiotics in neurodegenerative diseases: new hope for Parkinson's disease? *Neural Regen Res.* 2021 Apr;16(4):628–34.
- [13] Wong CB, Kobayashi Y, Xiao J. Probiotics for preventing cognitive impairment in Alzheimer's disease [Internet]. *Gut Microbiota - Brain Axis.* IntechOpen; 2018 [cited 2021 Sep 14]. Available from: <https://www.intechopen.com/chapters/62125>
- [14] Gandhi KR, Saadabadi A. Levodopa (L-Dopa). In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 [cited 2021 Sep 14]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK482140/>
- [15] Lewitt PA. Levodopa for the treatment of Parkinson's disease. *N Engl J Med.* 2008 Dec 4;359(23):2468–76.
- [16] Treur J. Network-oriented modeling: addressing complexity of cognitive, affective and social interactions [Internet]. Springer International Publishing; 2016 [cited 2021 Sep 14]. (Understanding Complex Systems). Available from: <https://www.springer.com/gp/book/9783319452111>
- [17] Treur J. Network-oriented modeling for adaptive networks: designing higher-order adaptive biological, mental and social network models [Internet]. Springer International Publishing; 2020 [cited 2021 Sep 14]. (Studies in Systems, Decision and Control). Available from: <https://www.springer.com/gp/book/9783030314446>

# Studies on the Decrease Mechanisms of Typical Complex Networks

Yuhu QIU<sup>a</sup>, Tianyang LYU<sup>b,1</sup>, Xizhe ZHANG<sup>c,d</sup>, Ruozhou WANG<sup>d</sup>

<sup>a</sup>*University of International Relations, Beijing 100091, China*

<sup>b</sup>*IT Center, National Audit Office, Beijing 100073, China*

<sup>c</sup>*School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing 211166, China*

<sup>d</sup>*School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China*

**Abstract.** Network decrease caused by the removal of nodes is an important evolution process that is paralleled with network growth. However, many complex network models usually lacked a sound decrease mechanism. Thus, they failed to capture how to cope with decreases in real life. The paper proposed decrease mechanisms for three typical types of networks, including the ER networks, the WS small-world networks and the BA scale-free networks. The proposed mechanisms maintained their key features in continuous and independent decrease processes, such as the random connections of ER networks, the long-range connections based on nearest-coupled network of WS networks and the tendency connections and the scale-free feature of BA networks. Experimental results showed that these mechanisms also maintained other topology characteristics including the degree distribution, clustering coefficient, average length of shortest-paths and diameter during decreases. Our studies also showed that it was quite difficult to find an efficient decrease mechanism for BA networks to withstand the continuous attacks at the high-degree nodes, because of the unequal status of nodes.

**Keywords.** Complex network, decrease mechanism, network evolution

## 1. Introduction

The running of complex systems like human society, biosphere rely on the interactions among large number of objects that form those systems. Complex network emerged as an important tool to study the interactions among objects since 1990s [1, 2]. It has been widely applied in sociology [3], physics [4], biology [5], virology [6], information science [7] and etc.

Researchers proposed complex network models to simulate the interactions and evolutionary features of lots of objects/nodes. As one of the earliest network model, ER model supposed that nodes interacted with each other randomly [8]. Then, dynamic models with evolutionary features were proposed. Barabasi and Albert [9] proposed the scale-free models that embedded the preferential attachment rule for growth and simulated the rich-get-richer phenomenon in the degree distribution of complex networks.

---

<sup>1</sup> Corresponding author; Tianyang LYU, IT Center, National Audit Office, Beijing 100073, China; E-mail: raynor1979@163.com.

These have become the foundations to understand, control and manipulate complex networks [6].

Network decreases referred to the removal, disappearance or deletion of nodes [10] that may lead to massive decrease of a network. In network evolution, decrease is a process that has the same importance as growth. A growth mechanism enabled a network expanded in suitable cases, while a decrease mechanism helped a network survive in harsh time [11]. Decreases have happened frequently in our normal life [11, 12], such as the decapitation strike on key individuals in a terrorist network, the financial crisis caused by the bankruptcy of Lehman Brothers, the failure of some nodes that lead to power grid breakdown in Europe and American. Previous studies [11-12] have figured out the influences of node removal and these influences varied with networks. Albert *et al.* [11] showed that random failures of nodes in random connection networks had accumulated consequences, but this was not the case for scale-free networks. Cohen *et al.* [12] showed that scale-free networks were vulnerable for attacks on high-degree nodes, and the weakness strengthened with the decrease of a network's power exponent. Hobbs and Burke [13] showed that attacks on specific nodes caused a cascade failure.

Complex networks in real life usually have a kind of coping strategy with network decreases, including replacement of a dying node, adjustments of interactions. For instance, terrorists always rebuilt their organizations after decapitation, and the social network also recovered from the death of an individual [10]. Therefore, it was critical for network models to contain a decrease mechanism. In the paper, a decrease mechanism referred to a strategy maintaining the existing network function with low cost when some nodes disappeared or even a network experienced a massive decrease.

However, many typical network models lacked a decrease mechanism. Take the scale-free model as an example. Although some experts proposed the decrease mechanism like preferential depletion [14], the mechanism was based on a fully connected network instead of a scale-free network. Bauke *et al.* [15] proposed to delete some nodes in network growth. In this case, decrease was still a by-product of growth, instead of an independent process that may lead to net reduction of network.

The paper proposed the corresponding independent decrease mechanisms for three typical models, including ER random network [8], WS small-world network [16] and BA scale-free network [9]. Each type of these networks has been widely found in the real life.

## 2. Decrease Mechanisms of Typical Network Models

A decrease mechanism should maintain topology features of a network. Topology features are usually critical for a network to perform its function and maintain its pattern. For instance, the existence of a few high-degree nodes greatly formed the propagation pattern of virus, information and action in a scale-free network, while the pattern was quite different in other types of networks [6]. Therefore, if a decrease made high-degree nodes vanish, the existed propagation mode would experience a big and costly change. From this perspective, the preferential attachment rule was sound that maintained the scale-free feature perfectly in growth. As a parallel mechanism, the decrease one should do the same job.

The key features of these typical networks that should be maintained in decrease were stated as follows. In general, the key feature of ER networks was random connections, this of WS networks was the long-range connection based on nearest-

coupled networks, these of BA networks were the preferential attachment rule and the scale-free property. The paper firstly explored the decrease mechanisms maintaining these features. Since the number of edges  $M$  of a certain number of nodes  $N$  greatly determined other metrics, the paper supposed that the average degree  $\bar{k}$  maintained before and after decrease of a network  $G^N$ .

As the removal of one node was the basis of network decreases, the paper proposed the decrease mechanisms of one node. It was quite simple to generalize the mechanisms to removal of lots of nodes.

### 2.1. The Decrease Mechanism of ER Networks

Nodes in an ER network  $G_{ER}$  connected with each other under the same probability  $P_{ER}^N = \frac{\bar{k}}{N-1}$ [8].

It was easy to prove that the removal of nodes in  $G_{ER}$  led to over deletion of edges among remain nodes. Eq. (1) showed the number of edges  $M'$  after remove  $\Delta N$  nodes. But an ER network with the same  $P_{ER}^N$  and  $N - \Delta N$  nodes should have  $M''$  edges as Eq. (2) showed. Therefore, the removal of  $\Delta N$  nodes results in over-deletion of  $\Delta M$  edges as Eq. (3) stated.

$$M' = \frac{N \times \bar{k}}{2} - \left( \Delta N \times \bar{k} - \frac{\Delta N \times (\Delta N - 1)}{2} \times P_{ER}^N \right) \quad (1)$$

$$M'' = \frac{(N - \Delta N) \times \bar{k}}{2} \quad (2)$$

$$\Delta M = \frac{\bar{k} \times (N_0 - \Delta N) \times \Delta N}{2 \times (N_0 - 1)} \quad (3)$$

$$\Delta M \xrightarrow{\Delta N=1} \frac{\bar{k}}{2} \quad (4)$$

This reflected the disadvantage of survivors after a continuous decrease of an ER network. Edges of these remaining nodes became more and more sparse. And the degree distribution of survivors changed so greatly that the left network should no longer be considered as an ER network.

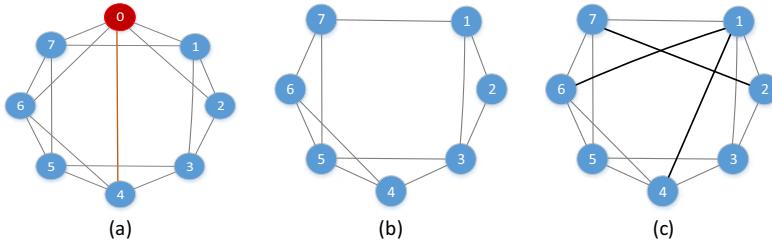
The proposed decrease mechanism of ER network was also simple. It randomly added  $\bar{k}/2$  edges among remaining nodes after remove one node as Eq. (4) shows. Then, the feature of ER network could maintain.

### 2.2. The Decrease Mechanism of WS Networks

A WS small-world network  $G_{ws}$  was constructed based on a nearest-coupled network by replacing an edge with the probability  $p_{ws}$ [16]. During the replacement, an edge  $e(i, j)$  replaced one of its end point  $v_i$  and with another node. Since new end point had a high probability to be a node far from the remaining one,  $G_{ws}$  showed small-world feature with short network diameter. And the replaced edge was usually called a long connection.

The higher the  $p_{ws}$ , the smaller the diameter. According to the feature of nearest-coupled network,  $\bar{k}$  was usually even.

There were two obvious influences of the removal of a node  $v_i$  in a WS network. First, the removal of  $v_i$  broke the nearest-couple relationships of  $v_i$ 's neighbors. Second, if  $v_i$  was the end point of some long-connections, vanish of many  $v_i$ s would influence the small-world feature of  $G_{ws}$ . Figure 1(a) and 1(b) showed an example, removal of  $v_0$  influenced the neighbor relation between  $v_0$  and  $v_6$  and the long-connection  $e(4,0)$ . To be more clearly, the paper placed nodes in a circle and the nearest neighbors of a node were placed at its both sides as Figure 1 showed.



**Figure 1.** A WS network with  $\bar{k}=4$  changed before and after the removal of  $v_0$ . (a) Original network (b) The network after  $v_0$  was deleted. (c)The network processed by the proposed decreased mechanism.

Therefore, the follow-up process was needed. Firstly, some edges of  $v_i$  should be replaced by other edges to maintain coupled relations of  $v_i$ 's nearest neighbors. Easy to prove that  $\bar{k}/2$  edges should be deleted and  $\bar{k}/2$  edges should be replaced. In the paper, if  $\bar{k}/2$  was even,  $\bar{k}/4$  edges at both sides of  $v_i$  were deleted and replaced respectively. If  $\bar{k}/2$  was odd,  $(\bar{k}+2)/4$  edges were removed and  $(\bar{k}-2)/4$  edges were replaced at the left side of  $v_i$ , and  $(\bar{k}-2)/4$  edges were removed and  $(\bar{k}+2)/4$  edges were replaced at the right side.

Secondly, handle the long-connections related with  $v_i$ . The strategy mainly focused on whether the original edge corresponding to a long-connection was planned to be deleted or replaced after decrease. The long-connection should be deleted if the original edge was, or the long-connection should be replaced.

Thus, we proposed the decrease mechanism of WS networks after the removal  $v_i$  as follows.

(1) Edge deletion.

- In the case of  $\bar{k}/2$  is even, delete all edges  $e(i,j)$ , if  $j \in \{i - \bar{k}/4, \dots, i + \bar{k}/4\}$  and  $i \neq j$ ;
- In the case of  $\bar{k}/2$  is odd, delete all edges  $e(i,j)$ , if  $j \in \{i - \frac{\bar{k}+2}{4}, \dots, i + \frac{\bar{k}-2}{4}\}$  and  $i \neq j$ ;

(2) Edge replacement.

- In the case of  $\bar{k}/2$  is even, replace all edges  $e(i,j)$ , if  $j$  satisfies the following requirements and  $i \neq j$  and  $e(i,j)$  is not replaced by a long-connection before decrease; when  $j \in \left\{i - \frac{\bar{k}}{2}, i - \frac{\bar{k}}{2} + 1 \dots i - \frac{\bar{k}}{4} - 1\right\}$ , replace  $e(i,j)$  by  $e(j + \frac{\bar{k}}{2} + 1, j)$ ; when  $j \in \left\{i + \frac{\bar{k}}{4} + 1, i + \frac{\bar{k}}{4} + 2 \dots i + \frac{\bar{k}}{2}\right\}$ , replace  $e(i,j)$  by  $e(j - \frac{\bar{k}}{2} - 1, j)$ ;

- In the case of  $\bar{k}/2$  is odd, replace all edges  $e(i,j)$ , if  $j$  satisfies the following requirements and  $i \neq j$  and  $e(i,j)$  is not replaced by a long-connection before decrease; when  $j \in \{i - \bar{k}/2, i - \bar{k}/2 + 1, i - (\bar{k} + 2)/4 - 1\}$ , replace  $e(i,j)$  by  $e(j + \frac{\bar{k}}{2} + 1, j)$ ; when  $j \in \{i + (\bar{k} - 2)/4 + 1, i + (\bar{k} - 2)/4 + 2, \dots, i + \bar{k}/2\}$ , replace  $e(i,j)$  by  $e(j - \frac{\bar{k}}{2} - 1, j)$ ;

(3) Handle long-connections.

- When the edge  $e(i,j)$  to be replaced in the above process is already replaced by a long-connection  $e(i,p)$  or  $v_i$  becomes an end point of a long-connection, replace  $e(i,p)$  by  $e(i-1,p)$  or  $e(i+1,p)$  randomly; when  $e(i,j)$  is replaced by a long-connection  $e(j,p)$ , removal of  $v_i$  has no influence on the long-connection and no more action needed.
- When  $v_i$  was a faraway node of  $e(i,p)$  after replaced the original edge  $e(q,p)$ , replace  $e(i,p)$  by  $e(i-1,p)$  or  $e(i+1,p)$  randomly.
- Randomly delete  $\Delta N \times \frac{\bar{k}}{2} \times p_{ws}$  long-connections to maintain  $p_{ws}$  during decrease, where  $\Delta N$  is the number of removal nodes.

Figure 1(c) showed an example. With  $\bar{k}=4$ ,  $e(0,7)$  and  $e(0,1)$  were deleted;  $e(0,6)$  was replaced by  $e(1,6)$  and  $e(0,2)$  was replaced by  $e(7,2)$ ; the long-connection  $e(0,4)$  was replaced by  $e(1,4)$ .

Using this mechanism, a WS network could maintain its nearest-coupled and small world features. But the decrease mechanism was much more complicated than that of ER. It had to perform complex edge replacements around a deleted node and keep tracking the status of at most  $N \times \frac{\bar{k}}{2} \times p_{ws}$  long-connections.

### 2.3. The Decrease Mechanism of BA Networks

The major feature of BA networks was rooted in the preferential attachment in network growth, which means new node tended to connect with high-degree nodes. This made a network evolved the scale-free property. Thus, the earlier a node joined a network, the higher its degree might be.

Therefore, the key of a decrease mechanism was to maintain the existence of high-degree nodes, especially the early nodes with high-degree. Easy to prove that, for a BA network  $G_{BA}$  with  $\bar{k}$ , the degree of a new node was  $\bar{k}/2$ . If the degree of a node  $v_j$  larger than  $\bar{k}/2$ , it was caused by connections added by nodes latter than  $v_j$ . Based on this analysis, the basic idea of the proposed mechanism was to delete edges connected at the moment when  $v_j$  joined the network, but rebuilt those connected after  $v_j$ . Thus, it looked like  $v_j$  has never existed when newcomers after  $v_j$  joined the network.

Therefore, the decrease mechanism of BA network was stated as follows,

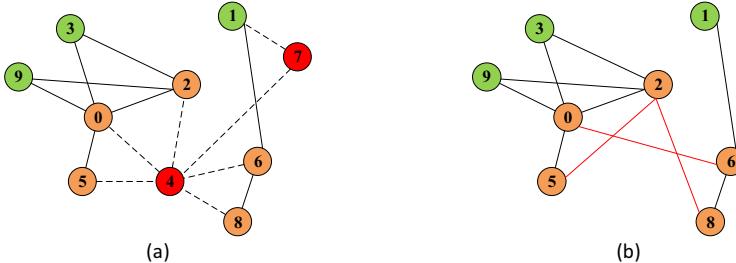
(1) In the case of the degree of  $v_j$   $k_j = \bar{k}/2$ , delete  $v_j$  and its edges.

(2) In the case of  $k_j > \bar{k}/2$ ,  $k_j - \bar{k}/2$  edges should be rebuilt. Let the set of nodes joined before  $v_j$  record as  $A_{j-}$  and the set of  $v_j$ 's neighbors that belong to  $A_{j-}$  record as  $B_{j-}$ ; let the set of nodes joined before  $v_j$  record as  $A_{j+}$  and the set of  $v_j$ 's neighbors that belong to  $A_{j+}$  record as  $B_{j+}$ . According to the growth mechanism of a BA network,  $|B_{j-}| = \bar{k}/2$  and  $|B_{j+}| = k_j - \bar{k}/2$ . The rebuilding rule is:

(a) Randomly select a node  $v_i$  from  $B_{j+}$ ;

- (b) Select a high-degree node  $v_k$  from  $A_{j-}$  according to the preferential attachment rule and no edge existing between  $v_k$  and  $v_i$ , rebuild  $e(k, i)$ ;  
 (c) Delete  $v_i$  from  $B_{j+}$ ;  
 (d) Repeat (a)-(c) until  $B_{j+}$  is empty.

Figure 2 showed an example where nodes were numbered according to their sequence joining the network and  $v_0$  was the first. When a newcomer  $v_7$  was deleted, only  $e(7,1)$  and  $e(7,4)$  were deleted. But when  $v_4$  was deleted, the edges created by adding  $v_5$ ,  $v_6$ ,  $v_7$  should be rebuilt. Thus,  $e(6,4)$  was replaced by  $e(6,0)$ ,  $e(5,4)$  was replaced by  $e(5,2)$  and  $e(8,4)$  was replaced by  $e(8,2)$ .



**Figure 2.** A BA network with  $\bar{k} = 4$  changes before and after the removal of  $v_4$  and  $v_7$  (a) Original network  
 (b) The rebuilt network after the removal of  $v_4$  and  $v_7$ .

This decrease mechanism maintained the local connectivity around a removal node and the high-degree property of first comers in a BA network. In comparison with those of ER and WS networks that have no growth mechanism like BA networks, the BA mechanism needed the time-dependent global information of the joining sequence of all nodes, while the mechanism of ER needed no additional information and that of WS needed only global information of a few long-connections. However, such global information was often rare for the Internet and social networks. Moreover, the performance of this mechanism to cope with continuous attacks on high-degree nodes still needed further verification.

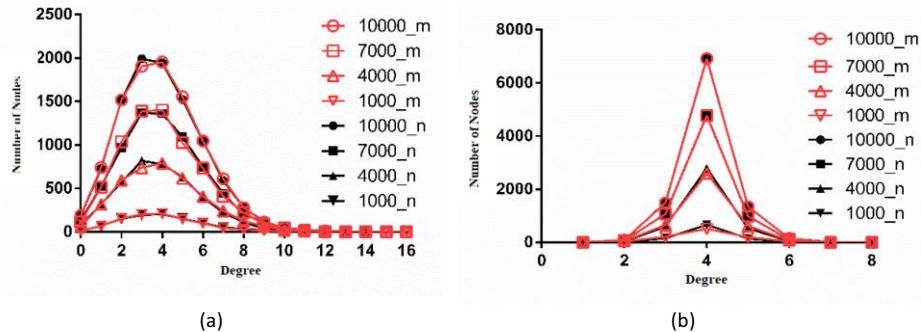
### 3. Experiments and Analysis

In experiments, we verified whether the proposed mechanisms maintained the topology characteristics like degree distribution, clustering coefficient, average length of shortest-paths and diameter during decrease, as they maintained the major network features.

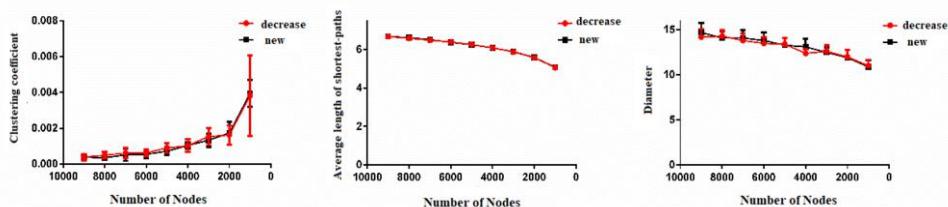
For each type of networks, we first generated 5 initial networks with  $N_0=10000$ ; all these networks decreased until  $N=1000$ . To avoid the influence of randomness, the decrease process was performed 20 times independently and the average of these metrics of the decreased networks were used in comparison with those of new created networks with the same size. If their metrics were similar, it means decreases caused no significant influence on a network. Let  $1000\_n$  correspond to a new network with  $N=1000$ , while  $1000\_m$  corresponded to a network decreased from initial to  $N=1000$ . For BA networks, we also analyzed their performances when there were continuous attacks on high-degree nodes. And  $1000\_d$  corresponded to a BA network decreased after target attacks.

### 3.1. Decrease Experiments of ER and WS Networks

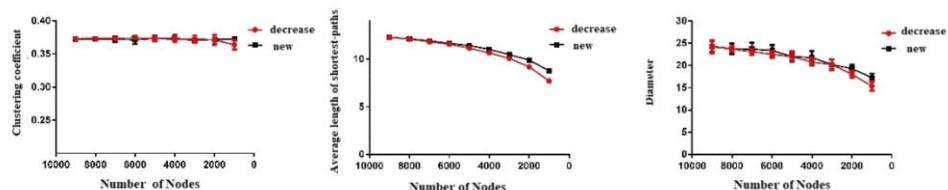
The proposed mechanisms of ER and WS networks performed well in the experiments. Figure 3, Figure 4 and Figure 5 showed that the metrics of new ER or WS networks were quite similar with those of the decreased networks, in the cases when the sizes decreased to 7000, 4000 and even 1000. The results also showed that the proposed mechanisms maintained the key features of ER and WS networks. Only WS networks showed a little difference in metrics including the clustering coefficient and the average length of shortest-paths when networks decreased to the small size in Figure 5. It may be caused by the influence of decreases on long-connections.



**Figure 3.** Degree distribution comparisons of ER and WS networks before and after decrease. (a) Degree distribution comparisons of ER networks. (b) Degree distribution comparisons of WS networks.



**Figure 4.** Topology metrics comparisons of ER networks with and without decreases.

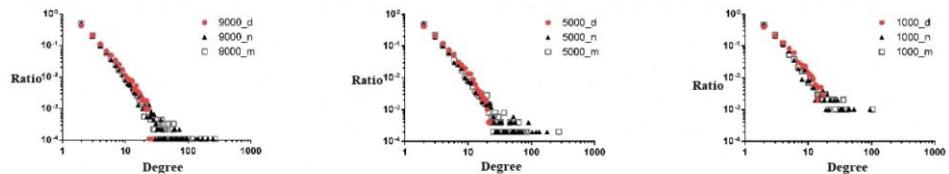


**Figure 5.** Topology metrics comparisons of WS networks with and without decreases.

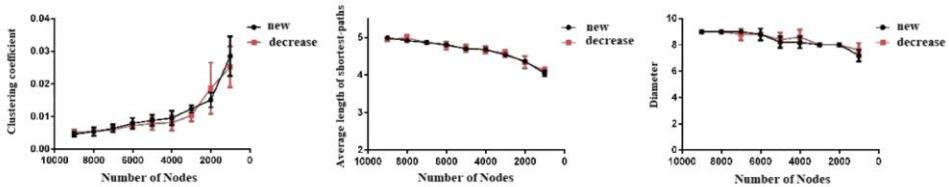
### 3.2. Decrease Experiments of BA Networks

Since there were clear differences of node status in a BA network, we analyzed the decrease performances after random deletion, and after target deletion of high-degree nodes. The results showed that the proposed mechanism handled the random deletion quite well. The metrics of new BA networks were quite similar with the networks decreased to 9000, 5000 and 1000 in Figure 6 and Figure 7.

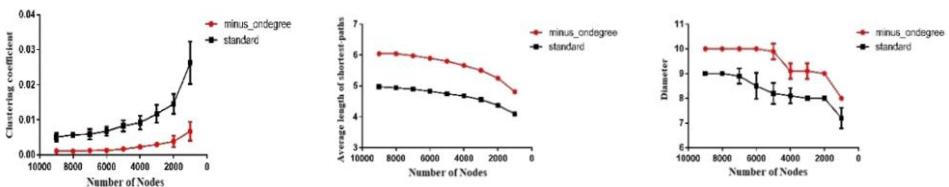
However, due to BA's intrinsic drawback, the mechanism was not good enough to encounter the continuous attacks on high-degree nodes. Figure 6 showed that the long tail vanished after the removal of 1000, 5000 and 9000 high-degree nodes. Figure 8 further showed the serious consequences on other metrics after continuous target attacks, where *minus\_ondegree* means target attacks and *standard* means new created BA networks. Preliminary analysis showed that the degree of target deleted nodes  $k >> \bar{k}/2$ . In this case, lots of edges were rebuilt, but the preferential attachment process cannot guarantee the existence of too many super star nodes. This led to a relative fair distribution of degree. Therefore, the reconstruction of all edges of a decreased network could solve this problem. However, the cost was too heavy to be applicable in real life.



**Figure 6.** Changes of the degree distribution after random and target deletion of nodes of BA networks.



**Figure 7.** Changes of the degree distribution after random deletion of nodes of BA networks.



**Figure 8.** Changes of topology properties after target deletion of nodes of BA networks.

## 4. Conclusion

The paper proposed the decrease mechanisms for three typical complex networks including ER, WS and BA. The study showed their good performance in maintaining the key features and topology characteristics of all these networks in decrease. However, we also found that some mechanisms also need local or even global information to handle the complex cases emerged after nodes' removal. From this perspective, decrease may be much more complex than growth that needs more attentions. Moreover, a sound decrease one for BA networks that can withstand continuous target attacks was still open for discussion, and detail observations of real networks on their decrease processes were critical for future studies.

## Acknowledgement

This work was partially supported by the Natural Science Foundation of China under grant number 91546110 and by the Scientific Research Funds for the Construction of High-grade, Precision and Advanced Discipline of University of International Relations under grant number 2019GA32.

## References

- [1] Barabási A, Crandall RE. Linked: the new science of networks. *Amer. J. Phys.* 2002; 71: 243-270.
- [2] Watts DJ. The "new" science of networks. *Annual Rev. Sociol.* 2004; 30:243.
- [3] Latora V, Marchiori M. Efficient behavior of small-world networks. *Phys. Rev. Lett.* 2001; 87:198701.
- [4] Albert R, Barabási A. Statistical mechanics of complex networks. *Rev. Modern Phys.* 2001; 74: xii.
- [5] Daura-Jorge FG, Cantor M, Ingram SN, Lusseau D, Simões-Lopes PC. The structure of a bottlenose dolphin society is coupled to a unique foraging cooperation with artisanal fishermen. *Biol. Lett.* 2012; 8: 702.
- [6] Pastor-Satorras R, Vespignan A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* 2001; 86 3200.
- [7] Maslov S, Sneppen K, Zaliznyak A. Pattern detection in complex networks: correlation profile of the internet. *Chromosoma.* 2002; 2002(3): 113-115.
- [8] Erdős P, Rényi A. On random graphs I. *Publicationes Mathematicae* 1959; 6: 290-297.
- [9] Barabási AL, Albert R. Emergence of scaling in random networks. *Science* 1999; 286: 509.
- [10] Motter AE, Lai YC. Topology of the conceptual network of language. *Phys. Rev. E Statistical Nonlinear Soft Matter Physics.* 2002; 66: 065102.
- [11] Albert R, Jeong H, Barabasi AL. Error and attack tolerance of complex networks. *Nature*, 2001; 340: 388-394.
- [12] Cohen R, Erez K, Ben-Avraham D. Breakdown of the internet under intentional attack. *Phys. Rev. Lett.* 2001; 86: 3682.
- [13] Hobbs WR, Burke MK. Connective recovery in social networks after the death of a friend. *Nature Human Behaviour.* 2017; 1: 0092.
- [14] Schneider CM, Arcangelis LD, Herrmann HJ. Scale free networks by preferential depletion. *EPL* 2011; 95: 1305-1323.
- [15] Bauke H, Moore C, Rouquier JB, Sherrington D. Topological phase transition in a network model with preferential attachment and node removal. *The European Phys. J. B.* 2011; 83(2011): 519.
- [16] Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature.* 1998; 393: 440.

# Application of Particle Swarm Optimization in Formulation Design of Pig Feed

Meijuan ZENG<sup>a,1</sup> and Hao LI<sup>a</sup>

<sup>a</sup> College of Automation, Wuhan University of Technology, Wuhan, Hubei, China

**Abstract.** Feed formula design is the core technology of accurate feeding management of livestock and poultry. High quality feed formula can not only improve the comprehensive and effective utilization of feed, give full play to the growth performance of animals, but also reduce production costs, so as to improve the economic benefits of farmers. Due to the characteristics of fast convergence speed and strong optimization ability of particle swarm optimization algorithm, this paper adopts particle swarm optimization algorithm to solve the pig feed formula, and builds the simulation model of the pig feed formula through Matlab to find the optimal solution for the feed formula. At the same time, compared with the traditional linear programming method, the simulation results show that, compared with the traditional linear programming method, the particle swarm optimization algorithm has better optimization results in solving the feed formulation problem.

**Keywords.** feed formulation; particle swarm optimization; linear programming

## 1. Introduction

Feed formula plays a very important role in the livestock and poultry breeding industry. High efficiency and high-quality feed formula can promote the healthy development of the livestock and poultry breeding industry. At present, the traditional manual calculation method has been difficult to meet the needs of actual production, only the computer application technology and feed formula fully combined, in order to better improve the feed conversion rate, shorten the growth cycle of live pigs, so as to reduce economic costs. The traditional computer optimization methods mainly include linear programming, objective programming and fuzzy programming. The lowest cost formula that meets certain constraints is calculated by decision [1]. However, this kind of method lacks flexibility and can only solve some problems with small operation scale, so it is difficult to properly deal with complex feed formulation problems such as multiple constraints or multiple objectives [2].

In recent years, genetic algorithm and ant colony algorithm have also been widely used in the design of feed formula. In reference [3], genetic algorithm was used to carry out feed matching, and the standard genetic algorithm was optimized and improved, so that the algorithm efficiency and formula benefit had achieved better results. Reference [4] adopted the linear programming method, standard genetic algorithm and NSGA-II

---

<sup>1</sup> Corresponding Author, College of Automation, Wuhan University of Technology, Wuhan, Hubei, China;  
E-mail: 1250362615@qq.com.

algorithm to carry out the matching of pig feed, respectively. The comparative analysis showed that the NSGA-II algorithm could not only reduce feed cost, but also effectively reduce ammonia emission, which was more suitable for multi-objective feed formulation problem. Reference [5] applied the continuous space ant colony optimization algorithm to the problem of feed formulation, and demonstrated the effectiveness and practicability of the optimization algorithm through the matching results.

Particle swarm optimization algorithm is widely used in function optimization, image recognition, neural network training and other fields because of its features such as easy implementation, few parameters, fast convergence speed and strong optimization ability [6]. In this paper, we will model and analyze the feed formula for pig feed, take feed production cost as the optimization goal, the nutritional standard required for pig growth period as the constraint, construct the mathematical model of pig feed formula optimization, and solve the feed formula using linear planning method and particle group algorithm respectively, and compare the analysis to show that the particle group algorithm has stronger solving ability.

## **2. Pig Feed and Formula Model**

### *2.1. Feed Raw Materials and Nutritional Standards*

Feed raw materials in the process of pig growth mainly provides the life-sustaining essential crude protein, amino acids, vitamins, minerals, total calcium, total phosphorus and other nutrients [7], but a single feed is not enough to meet the nutrition needs of pig growth, and according to the nutritional needs of different stages of pig growth will be reasonable matching of different raw materials, not only conducive to the growth and development of pigs, but also improve the utilization rate of feed.

**Table 1.** Nutritional content and price of feed raw materials (%).

In order to facilitate the research, the data of feed materials and their nutritional composition and the feeding standard data of pigs in this paper are all selected from the NRC standard of the United States [8-9], and the price of raw materials adopts the average price of a certain place. The main raw materials and their nutritional composition [10], maximum and minimum limits of usage and prices of pig feed formulations were shown in Table 1.

In this paper, the feed formulation of pigs at the growth stage of 50-80kg was taken as the research objective, and the nutritional standards of pigs at this stage were shown in Table 2.

**Table 2.** Nutritional standard for pig feed of 50~80kg.

Nutrient	Nutritional standards
Digestible energy (Mcal/kg)	3.4
Crude protein (%)	15.5
Arginine (%)	0.27
Histidine (%)	0.24
Isoleucine (%)	0.42
Leucine (%)	0.71
Lysine (%)	0.75
Methionine (%)	0.2
Phenylalanine (%)	0.44
Threonine (%)	0.51
Tryptophan (%)	0.14
Valine (%)	0.52
Calcium (%)	0.5
Phosphorus (%)	0.45

## 2.2. Feed Formula Model of Pig

Pig feed formula design is a typical combination optimization problem, is different kinds of feed raw materials in proportion to a reasonable match to meet the specific nutritional needs of pigs at different weight stages and the lowest economic cost Eq. [7].

If M kinds of feed raw materials and N kinds of nutritional requirements are provided, the matrix  $P_{ij}$  ( $1 \leq i \leq n$ ,  $1 \leq j \leq m$ ) of  $N^*M$  can be formed, as follows:

$$P = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} \quad (1)$$

Set variable Z as the economic cost of feed, then the mathematical model of pig feed formula is as follows:

Objective function:

$$\min Z = c_1x_1 + c_2x_2 + \dots + c_nx_n \quad (2)$$

In Eq. (2),  $c_i$  is the price of various raw materials, and variable  $x_i$  is the amount of various raw materials in the formula (in percentage).

The constraint conditions of feed on nutrient composition are expressed by inequality group:

$$\left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_m \geq b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_m \geq b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_m \geq b_n \\ x_i \geq 0(i=1,2,\cdots,n) \\ \sum_{i=1}^n x_i = 100 \end{array} \right. \quad (3)$$

In Eq. (3),  $a_{ij}$  is the hundredth component of the j-th nutrient contained in the i-th raw material;  $b_i$  is the content of each nutrient in the feeding standard.

### 3. Solution to Feed Formula Model

#### 3.1. Linear Programming

When the linear programming method is used to solve the pig feed formula, it converts the relevant elements and constraints in the feed formula into linear mathematical functions, and solves the minimum value of the function under certain constraints [11]. In order to facilitate the operation, this paper uses the Linprog function that comes with Matlab to solve the linear programming to optimize the related problems.

The format of the Linprog function is as follows:

$$[x, fval] = \text{lingprog}(c, A, b, Aeq, beq, vlb, vub) \quad (4)$$

In Eq. (4),  $x$  is the optimal solution;  $fval$  is the optimal value of the objective function;  $c$  is the coefficient of the objective function;  $A$  is the coefficient matrix of the inequality constraint;  $b$  is the constant term of the inequality constraint;  $Aeq$  is the coefficient matrix of the equality constraint;  $beq$  is the constant term of the equality constraint;  $vlb$  is the lower bound of the variable;  $vub$  is the upper bound of the variable.

#### 3.2. Particle Swarm Optimization

The basic idea of the particle swarm optimization algorithm originates from the research on the foraging behavior of bird flocks. Bird flocks adjust their search paths through their own experience and information sharing between the flocks, so that the flocks can find the optimal destination.

A kind of particle is used to simulate a bird in the search space, and the position of each particle corresponds to the potential solution of the optimization problem. In the search process, particles are described by position and velocity vectors, where the

position vector represents the possible solution of the problem, and the velocity vector represents the direction and magnitude of the particle position change. Evaluate the fitness value of each particle through the set objective function, and iterate continuously until the optimal solution is found. The mathematical model of particle swarm optimization algorithm is as follows:

Suppose N particles search for the optimal solution in the d-dimensional space, and the position vector of the i-th particle in the space can be expressed as:

$$x_i = (x_{i1}, x_{i2}, \dots, x_{id}), i=1,2,\dots,n$$

Its velocity vector can be expressed as:

$$v_i = (v_{i1}, v_{i2}, \dots, v_{id}), i=1,2,\dots,n$$

The optimal position currently searched by the ith particle is called the individual extreme value, which can be expressed as:

$$pbest_i = (p_{i1}, p_{i2}, \dots, p_{id}), i=1,2,\dots,n$$

The optimal position currently searched by the entire population is called the global extreme value, which can be expressed as:

$$gbest = (p_{g1}, p_{g2}, \dots, p_{gd})$$

Each particle updates its position and velocity according to these two extreme values. The update formula is as follows:

$$\begin{cases} v_{id} = w * v_{id} + c_1 r_1 (p_{id} - x_{id}) + c_2 r_2 (p_{gd} - x_{id}) \\ x_{id} = x_{id} + v_{id} \end{cases} \quad (5)$$

In Eq. (5), w is the inertia weight coefficient;  $c_1$  and  $c_2$  are individual learning factors and social learning factors, respectively;  $r_1$ ,  $r_2$  are random numbers in the interval [0,1]; In general, for the efficiency and accuracy of the search, the position and velocity vector of the particles will be restricted by the boundary:

$$x_{\min} \leq x_i \leq x_{\max}, \quad v_{\min} \leq v_i \leq v_{\max}$$

## 4. Simulation and Result Analysis

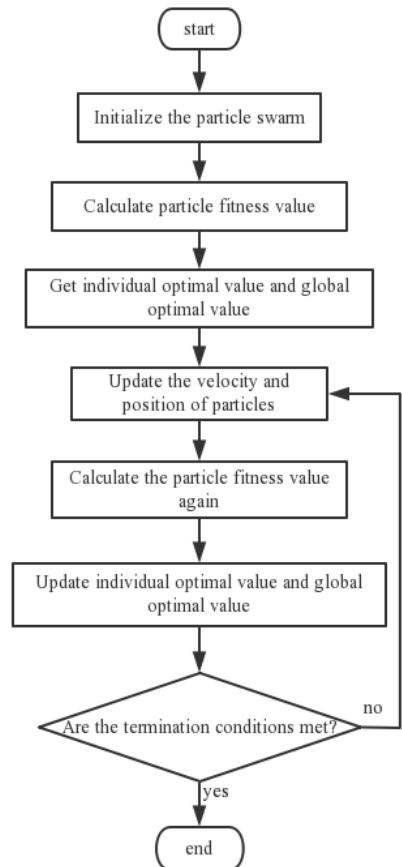
According to the feed ingredients and their nutrient content in Table 1 and the pig feeding standards in Table 2, 13 feed materials such as corn, wheat bran, soybean meal, fish meal, cottonseed cypress, etc. were selected to prepare 100kg pig feed.

### 4.1. Particle Swarm Optimization Simulation Experiment

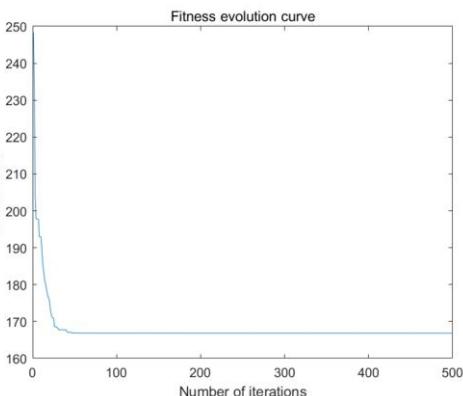
In the experiment, set the population size N=50, the maximum number of iterations G=500, the inertia weight w=0.9, and the learning factor  $c_1=2$ ,  $c_2=2$ . In the feed formula model established in this paper, the lowest economic cost is the objective function, and the fitness value takes the objective function value. The algorithm termination condition is: the particle swarm system loops 500 times.

The flow chart of particle swarm optimization algorithm is shown in the figure 1.

The optimization process of feed formulation by particle swarm optimization algorithm is shown in the figure 2, where the ordinate represents the optimal individual fitness value, and the abscissa represents the number of iterations. It can be seen from the convergence curve that the convergence speed of the algorithm slows down around 30 iterations, and stabilizes after 50 generations. This solution is the optimal solution for the pig feed formula. The result shows that the objective function value is 166.8732,  $x_1$ ,  $x_2$ , ...,  $x_{13}$  are 57.6771, 14.8748, 10.0000, 0, 1.5705, 7.0000, 0, 2.0000, 4.9216, 0, 0, 1.000, 0.3000 respectively.



**Figure 1.** Control flow chart of particle swarm optimization algorithm.



**Figure 2.** Feed formulation optimization process.

#### 4.2. Comparison and Analysis

Using particle swarm optimization algorithm and linear programming model respectively to solve the feed formulation, and the results are as follows, Table 3 and 4 are the output comparison of the two algorithms, the results show that, compared with linear programming, the cost of the feed formula obtained by particle swarm optimization algorithm is lower, and the nutrient elements in the formula are higher, also

fully meet the constraints of 25, that is, meet 14 constraints of nutritional requirements, and meet 11 constraints of upper and lower limits of usage.

**Table 3.** The cost comparison (kg).

Formula composition	Particle swarm optimization	Linear programming
Corn	57.6771	54.7172
Wheat bran	14.8748	20.0000
Soybean meal	10.0000	10.0000
Fish meal	0	0
Cottonseed meal	1.5705	0.0467
Rapeseed meal	7.0000	7.0000
Vegetable oil	0	0
Stone powder	2.0000	2.0000
Dicalcium Phosphate	4.9219	4.9361
Methionine	0	0
Lysine	0	0
Compound premix	1.0000	1.0000
Salt	0.3000	0.3000
total cost (yuan/100kg)	166.8732	167.2723

**Table 4.** The comparison of nutritional element standards (%).

Nutrient content	Particle swarm optimization	Linear programming	Standard requirement
Digestible energy (Mcal/kg)	3.4002	3.4000	3.40
Crude protein	16.3362	16.2604	15.50
Arginine	1.0549	1.0271	0.27
Histidine	0.4723	0.4695	0.24
Leucine	0.6105	0.6078	0.42
Isoleucine	1.3754	1.3590	0.71
Methionine	0.7503	0.7500	0.75
Lysine	0.2884	0.2855	0.20
Threonine	0.7571	0.7435	0.44
Tryptophan	0.6123	0.6091	0.51
Phenylalanine	0.1861	0.1881	0.14
Valine	0.7243	0.7241	0.52
Calcium	2.0193	1.9730	0.50
Phosphorus	1.1534	1.1385	0.45

## 5. Conclusions

Particle Swarm Optimization (PSO) has been applied more and more in the field of combinatorial optimization, but there is little research on solving the problem of pig feed formulation. In this paper, a new method for solving the problem of feed formulation is presented. In this paper, a new method for solving the problem of feed formulation is presented. The model of feed formulation for pigs is established to optimize the feed formulation. The results are compared with those obtained by linear programming algorithm under the unified model. The results show that the particle swarm optimization algorithm has better results for the optimization of pig feed formula, lower economic cost and higher nutritional level, thus proving the advanced and effective of the algorithm. In addition, according to the different nutritional requirements of the pigs at different weight stages, the feed formula optimization plan can be quickly and conveniently generated, which is conducive to shortening the growth cycle of the pigs and saving economic costs, thereby further achieving precise feeding. However, this paper mainly

takes economic cost as the goal. If we can combine the ammonia emission, lean meat ratio and other targets to carry out the feed ratio, a more optimized pig feed formula will be obtained.

## References

- [1] Zhang JX, Wu ZD. Application of particle swarm optimization technology in feed formulation design. Feed Industry 2010; 31(15): pp 47-48.
- [2] Wang JR, Wang TP. Mathematical model of pig feed formula optimization based on intelligent algorithm China Feed 2021; 8: 145.
- [3] Cao ZY. Application research of genetic algorithm in pig feed formula design (China: Yunnan University of Technology), 2016.
- [4] Huang K. Research on the application of genetic algorithm in pig feed formula system, China: Soochow University, 2014.
- [5] Feng J, Yue Q. An ant colony optimization algorithm for solving feed formula problem. J. Northeast Agricul. Uni. 2010; 41(03): 130-136.
- [6] Liu ZH, Wei HL, Zhong QC et al. Parameter estimation for VSI-Fed PMSM based on a dynamic PSO with learning strategies. In: IEEE Trans. Power Electr. 2017; 32(4): 3154-3165.
- [7] Yang SS. Research on optimized pig feed formula based on improved genetic algorithm (China: Anhui Agricultural University), 2019.
- [8] Meng QX et al. Dairy Cow Nutritional Needs 7th Revised Edition (2001) China: China Agricultural University Publishing, 2002.
- [9] Shixian Q et al. Nutritional Needs of Pigs (Tenth Revised Edition, 1998), China: China Agricultural University Publishing, 1998.
- [10] Huang K, Chang JY, Xie CH. Feed formula design based on NSGA-II Comp. Knowledge Technol. 2013; 9(20): 4692-4695.
- [11] Zhang YC. Application of mathematical methods in feed formulation research and neural network evaluation of results. China: Inner Mongolia University, 2019.

# A Comprehensive Energy Optimized Operation Strategy Based on Energy Storage Model

Bi-yan MA<sup>a 1</sup>, Xiao-dan ZHANG<sup>a</sup> and Hao-jie ZHOU<sup>a</sup>

<sup>a</sup> *Guangzhou Dongfang Dianke Automation Co. Ltd., Guangzhou 510000, China*

**Abstract.** The unified modeling of distributed resources in comprehensive energy parks is the basis to achieving energy scheduling. This article first analyzed the key modeling elements of the energy storage model, the conversion of central air conditioning and ice storage to the energy storage model was completed. Finally, based on the unified energy storage model in combination with the time-of-use electricity prices and load conditions of the park, the local economic operation strategy of distributed resources was realized. The energy storage model and local operation strategy were used as the edge computing of the energy management gateway, which realized the real-time status monitoring and economic operation scheduling of the comprehensive energies. This study provides local solutions for the economic operation of comprehensive energy parks.

**Keywords.** Energy storage model; ice storage; central air conditioning

## 1. Introduction

In recent years, the development of distributed energy has grown rapidly. The comprehensive energy system has a wealth of power load control methods such as energy storage, demand side response, and flexible load, and thus can quickly respond to the peak and frequency modulation services of the grid. However, when there is no grid dispatching curve, how to make comprehensive energy operate safely and economically locally according to the real-time operation of the grid, and maximize the benefits of multiple energy complements, has become an urgent issue to be solved [1-3]. Based on the study of the key elements of energy storage, ice storage and central air conditioning, this paper establishes a unified energy storage charging and discharging model, aiming to model the conversion of ice storage and central air conditioning to energy storage. Corresponding local economic operation strategies are formulated to realize the unified scheduling and the same local operation strategy implementation of a variety of different distributed resources, which provides flexible solutions for the control of local economic operation for parks with distributed resources.

---

<sup>1</sup> Corresponding Author, Bi-yan MA (1982), Guangzhou Dongfang Dianke Automation Co. Ltd., Guangzhou 510000, China; E-mail: 934412643@qq.com.

## 2. Energy Storage Model

The energy storage model is a model built for battery system-type energy storage charging and discharging stations. The main elements for modeling the battery system energy storage are:

Transformer capacity S. The PCS system in the energy storage station continuously monitors, using 0.95 as a reference value, the rated active power and maximum charging current of the energy storage station are calculated to be

$$P_0 = 0.95 * S \quad (1)$$

$$I_{max} = \frac{P_0}{3 * U_N} \quad (2)$$

where  $P_0$  is the rated active power of the energy storage station, S is the transformer capacity,  $U_N$  is the rated voltage of the transformer, and  $I_{max}$  is the maximum charging current of the energy storage station.

Total battery system capacity C. The total capacity of the battery system is a fixed parameter in the model.

Current SOC of the battery system. The battery system SOC is an important parameter of its energy storage status. The Equation of battery SOC at time t is

$$SOC(t) = SOC(t - \Delta t) \frac{\int_0^{\Delta t} i dt}{C} \quad (3)$$

where C is the total capacity of the battery system,  $SOC(t)$  is the SOC value of the battery at time t, and i is the operating current value of the battery at time  $(t - \Delta t)$ .

SOC threshold and SOH. To prevent the battery system from overcharging and overdischarging, thresholds for entering discharge  $SOC_{start\_discharge}$  and stopping discharge  $SOC_{discharge\_limit}$  are set. The equation for battery system SOH estimation is

$$SOH = \frac{SOC - SOC_{discharge\_limit}}{100 - SOC_{discharge\_limit}} \quad (4)$$

## 3. Ice Storage to Energy Storage Model

The ice storage air conditioner uses the low load power of power grid to make ice and store it in the ice storage device, and release the stored cold energy during the peak load period of the power grid to reduce the power load of air conditioning and the installed capacity of the air conditioning system during the peak time of the grid. The energy-storage-like behavior of ice storage can be modeled as an energy storage system:

### 3.1. Total Capacity of The Energy Storage System C

The maximum cold storage capacity of the ice storage system can be equivalent to the total capacity C of the energy storage system.

### 3.2. SOC Model

SOC of battery describes the state of charge, the SOC model conversion can be realized from the current remaining ice volume (SC\_left) and the total cold storage capacity (SC) of the ice storage system. The conversion Equation is as follows:

$$SOC = \frac{SC_{left}}{SC} \quad (5)$$

### 3.3. Maximum Charging Power P\_max

The maximum charging power refers to the maximum power that the current energy storage system can absorb from the grid. In the ice storage system, the power drawn from the grid can be composed of two parts, one is the power drawn by the cooling load, and the other is the power drawn by the ice making. Therefore, according to the demand of the cooling load and the current remaining ice volume, the following situations can be listed:

1. Current cooling load>0

- a) When current remaining ice capacity < cold storage capacity, the maximum charging power P\_max is the sum of ice making power and cooling power

$$P_{max}=P_{ice}+\min(P_{cool}, P_{QR})*F_{TR} \quad (6)$$

where P\_ice is the ice making power, P\_cool is the cooling power, P\_QR is the current cooling load, and F\_TR is the cooling ton conversion coefficient. 1 US cooling ton = 3024 kilocalories per hour (kcal/h) = 3.517 kilowatts (kw).

- b) Otherwise, when the current remaining ice capacity= cold storage capacity, the maximum charging power P\_max is just the cooling power

$$P_{max}=\min(P_{cool}, P_{QR})*F_{TR} \quad (7)$$

where P\_cool is the cooling capacity, P\_QR is the current cooling load, and F\_TR is the conversion coefficient of the cooling ton, as above.

2. No cooling load (e.g. no need for ice storage in winter)

There is no need for ice making and cooling, and therefore the maximum charging power P\_max is 0

$$P_{max}=0 \quad (8)$$

### 3.4. Maximum Discharge Power P\_max

The maximum discharge power refers to the maximum power that the current energy storage system can deliver to the grid. Since ice storage does not have the ability to transmit electrical energy, the current maximum reduction of power drawn from the grid can be equivalent to the maximum discharge power. The electricity consumption of the ice storage system mainly includes ice making and refrigeration. According to the current electricity consumption, they can be calculated as follows:

1. Remaining ice volume>0

Since ice storage has remaining ice, the cooling load can be provided completely by ice melting. Therefore, the current electricity used for ice making and cooling can be converted into discharge power. The calculation equation is as follows:

$$P_{\max} = Sts\_ice * P_{ice} + Sts\_cool * P_{cool} * F_{TR} \quad (9)$$

where  $P_{\max}$  is the discharge power,  $Sts\_ice$  is the state of the ice maker before discharging,  $P_{ice}$  is the ice making power,  $Sts\_cool$  is the cooling state before discharging,  $P_{cool}$  is the refrigeration capacity,  $F_{TR}$  is the conversion factor of the cooling ton, the same as above.

## 2. Remaining ice volume = 0

Since the remaining ice volume of the ice storage system is 0, the current cooling load cannot be provided by melting ice, that is, this part of the power cannot be reduced. Therefore, only the current electric power used for ice making can be converted into discharge power. The calculation equation is as follows:

$$P_{\max} = Sts\_ice * P_{ice} \quad (10)$$

where  $P_{\max}$  is the discharge power,  $Sts\_ice$  is the state of the ice maker before discharging, and  $P_{ice}$  is the ice making power.

## 4. Central Air Conditioning to Energy Storage Model

The air-conditioning system has a certain heat storage capacity, load adjustment can be carried out with fast response speed and the great potential. The indoor temperature range is recorded as  $[T_{min}, T_{max}]$ , when the room temperature of the air-conditioned room is less than  $T_{max}$ , turning off the air conditioner for a short time or increasing its set temperature can reduce the overall load level of the air-conditioning cluster [4-5]. Therefore, when the room temperature is lower than  $T_{max}$ , it can be considered that the air-conditioned room has stored part of the cooling capacity. Based on this idea, the energy storage model of a single air conditioner is established.

### 4.1. Total Capacity of Energy Storage System

The energy storage characteristics of central air conditioning reflect the load regulation ability conferred by the indoor temperature range. Therefore, the total energy storage capacity can be equivalent to the equivalent energy stored between the maximum room temperature and the minimum room temperature, as shown in Eq. (11):

$$\text{Capacity} = C(T_{max} - T_{min})/\eta \quad (11)$$

where Capacity is the total capacity of energy storage,  $C$  is the equivalent heat capacity of the air-conditioned room, indicating the amount of heat absorbed or released for every  $1^{\circ}\text{C}$  change in temperature. A value of  $1080\text{kJ}/^{\circ}\text{C}$  is taken in this model.  $\eta=3$  is the air-conditioning energy efficiency ratio.

### 4.2. SOC

$$\text{SOC equation is: } SOC = \frac{C(T_{max} - Tin)/\eta}{Capacity} = \frac{(T_{max} - Tin)}{(T_{max} - T_{min})} \quad (12)$$

where  $Tin$  is the current indoor temperature.

#### 4.3. Maximum Charge and Discharge Power $P_{max}$

##### 1. Maximum charge power

The rated power of the central air conditioning  $P_{max}$  can be equivalent to the maximum charging power  $P_{rate}$ , namely

$$P_{max} = P_{rate} \quad (13)$$

##### 2. Maximum discharge power

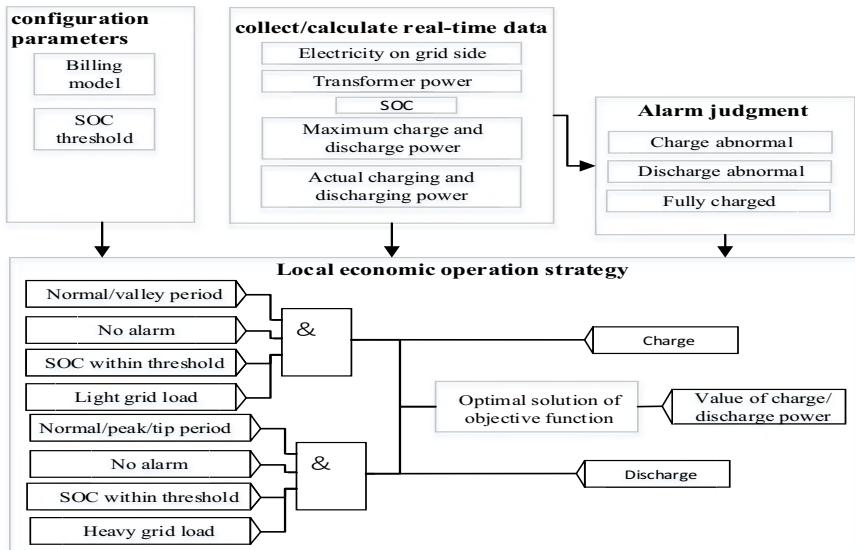
The discharge power  $P_{max}$  formula is as follows:

$$P_{max} = \frac{T_{out}(t) - T_{in}(t)}{\eta R} \quad (14)$$

where  $T_{out}(t)$  is the outdoor temperature, and  $T_{in}(t)$  is the indoor temperature.

#### 4.4. SOC Threshold

The maximum chargeable SOC ( $SOC_{max}$ ) and the minimum dischargeable SOC ( $SOC_{min}$ ) thresholds are defined, such that when the room temperature reaches the thresholds, the energy storage will stop, and normal operation starts (Figure 1).



**Figure 1.** Logic diagram of software framework.

### 5. Energy Storage Local Economic Operation Strategies and Model Simulation

#### 5.1. Energy Storage Local Economic Operation Strategies

After ice storage and central air conditioning are converted to energy storage model, energy storage, ice storage and central air conditioning can all share the same economic operation strategy to perform local optimization scheduling. Factors affecting the economic operation of energy storage mainly include time-of-use electricity prices,

current load of the park, etc. When the electricity price is in the valley, the energy is stored by turning on energy storage charging, starting ice making, and central air conditioning (or adjusting temperature). When it is at the peak price of electricity, the power drawn from the grid is reduced by discharging the energy storage, melting ice, turning off central air conditioning (or temperature adjustment), etc., thereby forming peak-to-valley arbitrage. When the power grid is in a normal period, the charging and discharging control needs to be dynamically determined according to the current park load situation, battery energy storage situation, and whether the next period type is peak period or valley period.

Device configuration parameters, equivalent calculated values of the energy storage model, and the alarm status of the energy storage model are used as the entry status of the local economic strategy. When there is no alarm in the valley period, enter the charge mode. When there is no alarm in the peak period, enter the discharge mode. Charging and discharging modes are flexibly switched during normal hours according to the current load level of the park. The charging and discharging power is solved by the optimal function method.

The energy storage model conversion algorithm is run in the distributed energy gateway as edge computing, and the charge state of various distributed energy sources is configured according to the load situation of the park. The feasibility of the algorithm is verified by running in real time. The distributed resources in the simulated park are as follows: Load 1 and Load 2 are energy storage warehouse loads, Load 3 is central air conditioning, and Load 4 is ice storage.

## 5.2. Model Conversion

### 5.2.1. Model Conversion of Ice Storage

Perform energy storage modeling on ice storage. The configuration parameters are shown in Table 1.

**Table 1.** configuration data of the ice storage model.

Name	desp	Unit	Meaning
ice_cool_storage	Cold storage capacity	Rth	
ice_cooling_capacity	Refrigeration capacity	Rt	Used to convert the maximum refrigerator power
ice_P_making_ice	Ice making power	kw	Used to convert the operating power of the ice maker
ice_max_melting	Maximum melting ice capacity	Rt	

**Table 2.** Real-time data of the ice storage model.

Name	Desp	Unit	Meaning
left_storage	Current remaining ice	Rth	
cooling_load	Current cooling load	Rt	Current demand of cooling capacity
ice_sts	Ice maker status		Current status: open/close
cooling_sts	Refrigerator status		Current status: open/close
cooling_output	Cooling capacity	Rt	
melting	Melted ice capacity	Rt	
soc	SOC	%	
max_power	Maximum charge/discharge power	kw	

Collect real-time data of ice storage, and calculate the equivalent SOC and maximum charge and discharge power in real time (Table 2). The simulation configuration data are entered, with the cold storage capacity, refrigeration capacity, ice making power, and maximum ice melting capacity being 60 Rth, 10 Rt, 15kw, and 10 Rt, respectively. The simulation real-time data are as follows: current remaining ice capacity 30 Rth, current cooling load 10 Rt, ice maker status: 1, and refrigerator status 1. The simulation enters the valley period, the SOC calculated by the gateway is 50%, and the maximum charging power is 50.17kw. The real-time values of the database are shown in Figure 2.

name	value	change_time_s	change_time_ms
load4_left_storage	30	0	0
load4_cooling_load	20	0	0
load4_eceduling_power	50.17	1616780870	313
load4_power	35.1	0	0
load4_ice_sts	1	0	0
load4_cooling_sts	1	0	0
load4_ice_ctrl	1	1616778245	202
load4_cooling_output	10	1616780878	206
load4_melting	10	1616780878	208
load4_soc	50	1616780865	770
load4_max_power	50.17	1616780865	774

Figure 2. Calculated value of load 4 during the valley period.

Notes: Database description: load4\_eceduling\_power is the charge and discharge power value of real-time scheduling. The value is positive for charging power, and negative for discharging power. The sum of the charging and discharging power of each distributed resource load should not be greater than maximum charging and discharging power value of the PCS. The maximum charge and discharge power of the park is configured to be 100kw. In this simulation, the real-time scheduling of ice storage charging power is 50.17kw, which does not exceed the maximum charging and discharging power of the park. Therefore, the real-time scheduling charge and discharge power value is 50.17kw, and the ice making is controlled to be turned on, and the load4\_ice\_ctrl value is 1. The “change\_time\_s” column in the database is the UTC timestamp when the database data is updated.

The simulated real-time data are as follows: current remaining ice volume 45 Rth, current cooling load 10 Rt, ice maker status 1, and refrigerator status 1.

The simulation enters the peak period. The SOC calculated by the gateway is 75%, the maximum discharge power is 50.17kw, and the real-time scheduling load discharge load4\_eceduling\_power is -50.17kW. And the control to shut down the ice making function load4\_ice\_ctrl value is 0. The real-time value of the database is shown in Figure 3.

name	value	change_time_s	change_time_ms
load4_left_storage	45	0	0
load4_cooling_load	10	0	0
load4_eceduling_power	-50.17	1616830214	241
load4_power	35.1	0	0
load4_ice_sts	1	0	0
load4_cooling_sts	1	0	0
load4_ice_ctrl	0	1616830266	131
load4_cooling_output	10	1616781956	147
load4_melting	0.000001	1616781956	149
load4_soc	75	1616830212	377
load4_max_power	50.17	1616830212	659

Figure 3. Calculated value of load 4 during peak period.

### 5.2.2. Central Air Conditioning Model Conversion

Perform energy storage modeling for central air conditioning. The configuration parameters are shown in Table 3:

**Table 3.** Central air conditioning model configuration data.

Name	Desp	Value	Unit	Meaning
AC_Capacity	Capacity		kJ	
AC_R	Equivalent thermal resistance of air-conditioned room	10.6	°C/kW	
AC_C	Equivalent heat capacity	1080	kJ/°C	
AC_eta	Air conditioning energy efficiency ratio	3		η
AC_Tmin	Minimum temperature	24	°C	
AC_Tmax	Maximum temperature	28	°C	
AC_SOCmax	Maximum charge SOC	90		Cannot charge after SOC reaches this value
AC_SOCmin	Minimum discharge SOC	10		Cannot discharge after SOC is less than this value
AC_P_rate	Rated power of air conditioner	3	kw	Use this power as the charging power when charging

The real-time data of central air conditioning were collected, and the equivalent SOC and maximum charge and discharge power in real time were calculated. The real-time data table is shown in Table 4.

**Table 4.** Real-time data of central air conditioning model.

Name	Desp	Unit	Meaning
soc	Remaining capacity		
Tin	Indoor temperature	°C	
Tout	Outdoor temperature	°C	
max_Power	Maximum charge/discharge power	Kw	
fault	Alarm status		

The simulation data are as follows: configure the cold storage capacity to be 136kJ, the equivalent thermal resistance of the air-conditioned room is 10.6°C/kW, and the equivalent thermal capacity is 1080 kJ/°C. The energy efficiency ratio of the air conditioner is 3, the minimum temperature is set to 24°C, and the maximum temperature is set to 28°C. The maximum charge SOC is 90%, the minimum discharge SOC is 10%, and the rated power of the air conditioner is 3kw. The simulated real-time data are as shown in Figure 4: indoor temperature 26.3°C, outdoor temperature 30°C. Simulation enters the valley period. The SOC calculated by the gateway is 42%, and the maximum charging power is 3kw (rated power of air conditioner).

name	value	change_time_s	change_time_ms
load3_soc	42	1616778224	760
load3_Tin	26.3	0	0
load3_Tout	30	0	0
load3_max_power	3	1616778245	171
load3_fault	0	0	0
load3_echeduling_power	3	1616778245	198

**Figure 4** Calculated value of load 3 during valley period.

Notes: Database description is the same as above. The simulated charging power load3\_echeduling\_power for real-time scheduling of central air conditioning is 3kw. The “change\_time\_s” column in the database is the UTC timestamp when the database data is updated.

The real-time database is as shown in Figure 5. Real-time data changes: Indoor temperature is 25°C, outdoor temperature is 30°C. The simulation enters the peak period, the SOC calculated by the gateway is 75%, and the maximum discharge power is 0.157233kw.

name	value	change_time_s	change_time_ms
load3_soc	75	1616824600	62
load3_Tin	25	0	0
load3_Tout	30	0	0
load3_max_power	0.157233	1616824600	65
load3_fault	0	0	0
load3_echeduling_power	-0.157233	1616824600	135

Figure 5. Calculated value of load 3 during peak period.

Notes: Database description is the same as above. The discharge power load3\_echeduling\_power for real-time scheduling of central air conditioning is -0.157233kw. Negative number represents discharge. The “change\_time\_s” column in the database is the UTC timestamp when the database data is updated.

### 5.3. Charge and Discharge Control

The billing model of Guangdong area is used. That is, “0-8 o’clock, valley period; 8-14 o’clock, normal period; 14-17 o’clock, peak period; 17-19 o’clock normal period, 19-22 o’clock, peak period; 22-0 o’clock, normal period”. Combining with the spontaneous or self-consumption situation of plants and stations, priority is given to ensuring self-consumption, followed by peak-to-valley arbitrage to realize charge and discharge control. The billing model database configuration is shown in Figure 6.

id	name	value
13	rate1_time	0
14	rate1_type	4
15	rate2_time	800
16	rate2_type	3
17	rate3_time	1200
18	rate3_type	2
19	rate4_time	1400
20	rate4_type	3
21	rate5_time	1500
22	rate5_type	2
23	rate6_time	1600
24	rate6_type	3

Figure 6. Billing model.

Notes: rate1\_time is the start time of billing period 1. Rate1\_type is the type of charging period 1, where 0 is the tip period, 1 is the peak period, 2 is the normal period, and 3 is the valley period, and so on for other periods.

Simulation enters the valley period, and the SOC is not full. After the gateway executes the program, it outputs the charging status and dispatches various distributed resource loads into the charging status. Figure 7 is a screenshot of the charge/discharge log of the device. The simulation enters the peak period, and the SOC is greater than the set value of termination of discharge. After the gateway executes the program, it outputs the discharge status and schedules various distributed resource loads to enter the discharge status. Figure 8 is a screenshot of the charge/discharge log of the device.

```
21/3/27 1: 4: 5 Start charging 1, soc=50.000000
21/3/27 1: 4: 5, Load participates in charging , 1, load_1
21/3/27 1: 4: 5, Load participates in charging , 1, load_2
21/3/27 1: 4: 5, Load participates in charging , 1, load_3
21/3/27 1: 4: 5, Load participates in charging , 1, load_4
```

Figure 7. Charging log during valley hours.

```
21/3/27 13: 56: 40 start discharging 1, soc=50.000000
21/3/27 13: 56: 40, Load participates in discharging , 1, load_1
21/3/27 13: 56: 40, Load participates in discharging , 1, load_2
21/3/27 13: 56: 40, Load participates in discharging , 1, load_3
21/3/27 13: 56: 40, Load participates in discharging , 1, load_4
```

Figure 8. Discharge log during peak hours.

## 6. Conclusion

This article first summarized the key elements of energy storage station modeling, and then analyzed the operating modes of ice storage and central air conditioning. Then the operating parameters of ice storage and central air conditioning were converted to the energy storage model elements to form a unified distributed energy model with energy storage as the main model. On the basis of unified modeling, according to the time-of-use electricity price, the current load of the park and other influencing factors, the charging and discharging control strategy of the local economic operation of the park's distributed resources was formulated. The model of energy storage, ice storage, central air conditioning, and local operation strategies were deployed as edge computing applications to the energy management gateway. After simulation calculation, the following conclusions are drawn: the energy storage model of ice storage and central air conditioning can be converted in real time according to the load condition. The energy management gateway used a unified model to conveniently realize the flexible scheduling without differentiation of multiple distributed resources. It provides local solutions for the economic operation of comprehensive energy parks, improves the economic benefits of energy parks, which makes it worth promoting.

## Reference:

- [1] Zhang YN, He YB, Yan MY, et al. Optimal dispatch of integrated electricity-natural gas system considering demand response and dynamic natural gas flow. Automation Electric Power Syst. 2018; 42(20): 1-10.
- [2] Wu M, Li ZW, Sun LJ. A model predictive overall control method for hybrid energy storage converter. Power Syst. Protection Contr. 2020; 48(21): 84-91.
- [3] Wang HH, Chen JJ, Zhu T et al. Joint bidding model and algorithm of wind-storage system considering energy storage life and frequency regulation performance. Power Syst. Technol. 2021; 45(01): 208-217.
- [4] Wang SX, Zhu W. Optimal scheduling of CCHP micro-grid with ice storage air-conditioning. J. Shanghai Univer. Electr. Power, 2021; 37(1): 37-43.
- [5] Cheng S, Huang TL, Wai RJ. Multi-time-scale optimal scheduling of CCHP microgrid with ice-storage air-conditioning. Automation Electr. Power Syst. 2019; 43(05): 30-38.

# Soil CEC Predicting Model of Tobacco-Planting Fields in Chenzhou, South Hunan Province

Y S XIAO<sup>a</sup>, Z H CAO<sup>a</sup>, W J LI<sup>a</sup>, Y H LIAO<sup>b</sup>,

B HE<sup>a</sup>, L J LI<sup>a</sup>, Z H HE<sup>a</sup>, Q ZHONG<sup>c, 1</sup>

<sup>a</sup>Chenzhou Tobacco Company of Hunan Province, Chenzhou, Hunan 423000, China

<sup>b</sup>Hunan Provincial Tobacco Company, Changsha 410010, China

<sup>c</sup>Yongxing Branch of Chenzhou Tobacco Company, Yongxing Hunan 423300, China

**Abstract.** [Background] Cation exchange capacity (CEC) is a basic but important soil property of soil fertility or quality, CEC predicting model is often derived from other soil properties measured more easily because the traditional method determining CEC is time-consuming and laborious. It is necessary to establish a new CEC prediction model for a new region because CEC predicting model usually is dependent on the study region. [Objective] Chenzhou City is the most important and typical tobacco-planting region with tobacco-rice rotation in Hunan province and China, this study was conducted to establish CEC predicting model for the tobacco-planting fields in Chenzhou because so far no CEC predicting model is available for tobacco-planting fields in Chenzhou and in China. [Method] In total 1055 topsoil samples (0~20 cm) were collected in 2015 from the tobacco-planting fields in Chenzhou, soil properties included the particle size composition, pH, soil organic matter and various nutrients were determined, the status of CEC were assessed, and then CEC predicting models were setup in different regions in Chenzhou. [Result] The results showed that CEC in Chenzhou was ranged from 3.50 to 48.50 cmol (+) kg<sup>-1</sup> with a mean of 22.05 cmol (+) kg<sup>-1</sup>, averagely belonged to the very high grade (>20 cmol(+)) kg<sup>-1</sup>). There were significant differences in CECs in different regions in Chenzhou, which was the highest in Jiahe (23.83 cmol(+)) kg<sup>-1</sup>) but the lowest in Anren (15.78 cmol(+)) kg<sup>-1</sup>). CEC was significantly correlated with different soil properties in different regions, which was significantly correlated with coarse sands, fine sands, clays, pH and total P in Chenzhou ( $R=0.312^{**} \sim 0.445^{**}$ ), significantly correlated with coarse sands, silts, fine sands, clays, pH, total P, exchangeable Ca<sup>2+</sup>, Mg<sup>2+</sup> and available Zn in Suxian ( $R=0.430^{**} \sim 0.684^{**}$ ), significantly correlated with coarse sands, fine sands, silts, clays, pH, total P, available B and Cu in Yongxing ( $R=0.321^{**} \sim 0.605^{**}$ ), significantly correlated with coarse sands, fine sands and clays and total P in Guiyang ( $R=0.330^{**} \sim 0.477^{**}$ ), significantly correlated with coarse sands, silts and total K in Yizhang ( $R=0.326^{**} \sim 0.466^{**}$ ), and only significantly correlated with fine sands in Jiahe ( $R=0.350^{**}$ ). The accuracy of CEC predicting model usually was lower when less properties involved. Based on the comparison of the R<sup>2</sup> and RMSE of the established CEC predicting models, it is recommended that the total model for Chenzhou could be used for Guiyang, Jiahe and Yizhang, while the regional models should be selected for Yongxing, Anren and Suxian. [Conclusion] This study proves further that different soil properties were most important for CEC predicting models in different regions, new CEC predicting models must be setup for a new study region, and soil organic matter is

<sup>1</sup> Corresponding author: Yongxing Branch of Chenzhou Tobacco Company, Yongxing Hunan 423300, China; E-mail: zhql126@126.com.

not a variable in soil CEC predicting models for tobacco-planting fields in Chenzhou, which are different from some previous studies.

**Keywords.** soil CEC, statistic information, correlation, predicting model

## 1. Introduction

Cation exchange capacity (CEC) is a basic soil property often used as an index of soil fertility or quality, understanding CEC plays important roles in guiding reasonable fertilization and soil improvement [1-2]. Meanwhile, CEC is also an index for soil classification or taxonomy [3-4]. Because the traditional methods determining CEC are time-consuming and laborious, more studies were conducted to setup CEC predicting model from soil properties measured more easily [5-14], and the results showed that for different regions or soils, the variables used in predicting soil CEC are different, for examples, Rahal and Alhumairi [5] predicted soil CEC in mid-Mesopotamian plain by using texture class, bulk density, total available water content, soil color, sodium adsorption ratio, electrical conductivity and Ca<sup>2+</sup>. Khaledian *et al.* [6] proved that soil CEC was affected by different variables in different situations, clay (positively correlated) and sand (negatively correlated) were the most influential variables for predicting CEC, CEC was significantly and negatively correlated with pH in agricultural land uses in Spain, significant positive relationship between CEC and OC in Spain, the USA, Iraq, and pasture. Shiri *et al.* [7] used the contents of silts, clays, sands, organic carbon and pH in modeling soil CEC in Iran. Seyedmohammadi *et al.* [8] proved soil organic carbon and clay could be used as input variables (positively correlated) for predicting CEC of paddy soils in Guilan province, northern Iran. Liao *et al.* [9] modeled soil CEC with organic matter, silt, clay and pH (positively correlated) as well as sand (negatively correlated) in Qingdao in China. Obalum *et al.* [10] found that CEC of coarse-textured soils in southeastern Nigeria increased with decreasing coarse sand but with increasing fine sand, silt correlated negatively with the CEC, clay and organic matter generally impacted positively on the CEC, and the best-fitting linear CEC function was attained with fine sand, clay, and organic matter. Yukseken and Kaya [11] predicted soil CEC in Hawaii by using organic matter and clay fraction (positively correlated) with other variables (specific surface area, activity, Atterberg limits, plastic, shrinkage, and modified free swell index). Seybold *et al.* [12] used organic matter and clay content (positively correlated) and pH (positively or negatively correlated) as the main variables to model soil CEC in USA. Krogh *et al.* [13] found that CEC of Danish soils could be modelled with clay and organic matter content (positively correlated), while silt and pH (positively correlated) might also contribute as predictor variables. Manrique *et al.* [14] found that clay, organic carbon (positively correlated) and pH (negatively correlated) could be used in predicting CECs for all soils, while clay and organic carbon used in predicting CECs of Alfisols, Inceptisols, Mollisols, Vertisols, Entisols, Spodosols, Spodosols.

For tobacco-planting soils in China, CEC is often measured and used as an index of soil fertility [15-23], and the relationship between CEC and other properties were also discussed in some studies [24-27]. Furthermore, some studies found that CEC is closely related to the chemical components of tobacco leaves (total sugar, reducing sugar, salt and nicotine etc.) [28-29], and high CEC is conducive to reducing the occurrence and harm of bacterial wilt and red weed diseases of tobacco [30-31].

Chenzhou City of south Hunan Province, with a long history of tobacco-planting as early as in 1593 and where most paddy fields are under tobacco-rice rotation [32], is the most important and typical planting region of flue-cured tobacco with burnt-pure sweet aroma in China [33]. The area of tobacco-planting in Chenzhou is about  $26.7 \times 10^3$  hm<sup>2</sup> in recent years, which plays an important role in ensuring the supply of high-quality tobacco leaves and the sustainable development of regional society and economy. Some literatures were published about tobacco-planting soil characteristics in Chenzhou [34-37]. Nowadays, a new round of tobacco-planting soil improvement is underway in Chenzhou and in other regions of China, it is helpful to understand further soil CEC in tobacco-planting fields in providing scientific instruction for this work, However, so far little information is available on soil CEC predicting model for tobacco-planting fields in Chenzhou and China, thus, in this study, the status of soil CEC in tobacco-planting fields in Chenzhou were studied and CEC predicting models were setup based on other soil properties.

## 2. Methods and Materials

### 2.1. Sources of Soil Data

The data of soil properties used in this study came from the tobacco-planting soil surveys conducted in 2015, which included 1055 topsoil samples (0~20 cm) collected from the typical tobacco-planting fields in different regions of Chenzhou.

The typical field was decided according to the spatial distribution uniformity of the tobacco-planting field, in each typical field the topsoil sample was collected randomly at 5~8 points with stainless steel drill and then were mixed completely. The measured soil properties were included particle composition, pH (H<sub>2</sub>O), organic matter (OM), total nitrogen (TN), phosphorous (TP) and potassium (TK), available nitrogen (AN), phosphorous (AP) and potassium (AK), exchangeable calcium (Ca<sup>2+</sup>) and magnesium (Mg<sup>2+</sup>), available sulfur (S), water-soluble chlorine (Cl), and available boron (B), iron (Fe), manganese (Mn), copper (Cu), zinc (Zn) and molybdenum (Mo). The detailed determination methods for soil properties could be found in related literatures [38-41].

### 2.2. Grading Standard of Soil CEC

No information was available in China on the grading standard of soil CEC for tobacco-planting fields, in this study, soil CEC in Chenzhou was divided into 5 grades as shown in Table 1 which was based on soil CEC classification of the 2<sup>nd</sup> national soil survey [42-43].

**Table 1.** Grading standard of soil CEC for tobacco-planting field in Chenzhou.

Grade	Very low	Low	Middle	High	Very high
Value cmol(+) kg <sup>-1</sup>	<6.2	6.2~10.5	10.5~15.4	15.4~20	≥20

### 2.3. Data Processing and Statistics

Microsoft Excel 2016 and IBM Statistics SPSS 22.0 software were used for statistical analysis of the data, and Duncan test method (2-tailed) was used for variance analyses and multiple comparisons [43-44].

## 3. Results

### 3.1. Statistics and Comparison of CEC

Table 2 shows the statistical results of CEC. CEC was ranged from 3.50 to 48.50 cmol(+) kg<sup>-1</sup> with a mean of 22.05 cmol(+) kg<sup>-1</sup>, which covered the whole 5 grades, but averagely belonged to the very high grade (>20 cmol(+) kg<sup>-1</sup>). Meanwhile, for total samples, CEC was in the moderate middle variation (C.V. =61.08% < 100%), very positive skewness distribution (Skewness>0.3) and flat peak distribution (Kurtosis≈0) [44-45].

CEC was the highest in Yongxing (24.11 cmol(+) kg<sup>-1</sup>) but the lowest in Anren (15.31 cmol(+) kg<sup>-1</sup>). CEC of Yongxing was not significantly higher than those of Jiahe and Guiyang, but significantly higher than those of other regions. CEC of Jiahe was not significantly higher than those of Guiyang and Beihu, but significantly higher than those of Yizhang, Suxian, Linwu and Anren. CEC of Guiyang was not significantly higher than that of Beihu, but significantly higher than those of Yizhang, Suxian, Linwu and Anren, CEC of Beihu was not significantly higher than that of Yizhang, but significantly higher than those of Suxian, Linwu and Anren, CEC of Yizhang was significantly higher than those of Suxian, Linwu and Anren, while no significant difference was found among Suxian, Linwu and Anren.

**Table 2.** Statistic information of tobacco-planting soil CECs in Chenzhou.

Region	Sample no.	Minimum cmol (+) kg <sup>-1</sup>	Maximum Cmol (+) kg <sup>-1</sup>	Mean±S.D. Cmol (+) kg <sup>-1</sup>	C.V. (%)	Skewness	Kurtosis
Total	1055	3.5	48.5	22.05±7.70	34.9	0.45	0.01
Guiyang	560	3.5	48.5	23.80±7.62ace	32.02	0.42	0.01
Yongxing	115	6.3	45	24.11±8.17a	33.88	-0.01	-0.15
Jiahe	110	9.3	40.4	23.83±6.32ac	26.51	0.41	0
Anren	100	5.8	30.7	15.31±5.15k	33.62	0.4	-0.17
Yizhang	96	6.2	37.6	18.10±5.77dfgi	31.86	0.8	0.53
Suxian	45	6.2	27.6	16.30±4.92hjk	30.2	-0.05	-0.16
Beihu	17	14.7	29.1	20.88±3.94bcieg	18.88	0.44	0.18
Linwu	12	8.8	28.6	15.78±5.70k	36.16	1.22	1.24

Note: values in the same column followed by different lowercase letters are significantly different at the 0.05 level.

### 3.2. Statistics and Comparison of Soil CEC in Each Grade

Table 3 shows the statistical results of the numbers and proportions of tobacco-planting fields in different grades of CEC. For all the samples in Chenzhou and the samples in Guiyang, Yongxing, Jiahe and Beihu, the sample proportions were all in the order of

the very high grade > the high grade > the middle grade > the low grade > very low grade, and the sample proportion of the very high grade was 57.25%~74.78%. However, for the samples in Anren, the sample proportion was in the order of the middle grade (32.00%) > the high grade (29.00%) > the low grade (20.00%) > the very high grade (18.00%) > the very low grade (1.00%), while for the samples in Yizhang and Suxian, the sample proportions were both in the order of the high grade (34.38% and 35.56%) > the middle grade (32.29% and 28.89%) > the very high grade (29.17% and 24.44%) > the low grade (4.17% and 11.11%), and no sample in the very low grade.

**Table 3.** Sample numbers and proportions in different grades of CEC.

Region	Very high ( $\geq 20$ )		High (15.4~20)		Middle (10.5~15.4)		Low (6.2~10.5)		Very low (<6.2)	
	Sample no.	%	Sample no.	%	Sample no.	%	Sample no.	%	Sample no.	%
Total	604	57.25	239	22.66	160	15.17	47	4.45	5	0.47
Guizhou	367	65.54	122	21.79	60	10.71	7	1.25	4	0.71
Yongxing	86	74.78	10	8.70	10	8.70	9	7.83	0	0
Jiahe	81	73.64	21	19.09	7	6.36	1	0.91	0	0
Anren	18	18.00	29	29.00	32	32.00	20	20.00	1	1.00
Yizhang	28	29.17	33	34.38	31	32.29	4	4.17	0	0
Suxian	11	24.44	16	35.56	13	28.89	5	11.11	0	0
Beihu	11	64.71	4	23.53	2	11.76	0	0	0	0
Linwu	2	16.67	4	33.33	5	41.67	1	8.33	0	0

### 3.3. Correlation between CEC and Other Properties

Table 4 shows the statistical results of other soil properties while Table 5 shows the Pearson correlation coefficients between CEC and other properties. It can be seen from Table 5 that CEC was significantly correlated with different properties in different regions, for example, CEC was significantly correlated with coarse sands, fine sands, clays and TP for all samples in Chenzhou ( $R=0.312^{**} \sim 0.445^{**}$ , in this paper only soil properties with  $R \geq 0.3^{**}$  was used to setup CEC predicting model, because this value usually means significant correlation existed [44-45]), significantly correlated with coarse sands, fine sands, silts, clays, pH, TP,  $Ca^{2+}$ ,  $Mg^{2+}$  and Zn for samples in Suxian ( $R=0.430^{**} \sim 0.684^{**}$ ), significantly correlated with coarse sands, fine sands, silts, clays and pH, TP, B and Cu for samples in Yongxing ( $R=0.321^{**} \sim 0.605^{**}$ ), significantly correlated with coarse sands, fine sands and clays and TP for samples in Guiyang ( $R=-0.330^{**} \sim 0.477^{**}$ ), and only significantly correlated with fine sands for samples in Jiahe ( $R=0.350^{**}$ ).

According to the significant correlation existed between CEC and other properties, the optimal regression models of CEC were established for Chenzhou and for different regions in Chenzhou (see Table 6). If judged from  $R^2$  and RMSE/S.D., the accuracy was higher for soil CEC predicting models in Suxian, Anren and Yongxing ( $R^2=0.795^{**}$ ,  $0.602^{**}$  and  $0.489^{**}$ , and RMSE/S.D.=0.57, 0.66 and 0.74, respectively), but the accuracy was lower for soil CEC predicting models in Jiahe, Guiyang and Yizhang ( $R^2=0.123$ , 0.288 and 0.231 and RMSE/S.D.=0.94, 0.85 and 0.89, respectively). Thus, it is recommended that for Yongxing, Anren and Suxian, their own regional models

should be used to predict soil CEC, but for Guiyang, Jiahe and Yizhang, the total model could be considered to predict soil CEC.

**Table 4.** Statistic information of other soil properties in Chenzhou (n=1055).

Soil property	Minimum	Maximum	Mean±S.D.	C.V.	Skewness	Kurtosis
Coarse sands	0	56	8±5	61.08	2.89	15.51
Fine sands	4	64	25±8	31.87	1.10	2.18
Silts	10	56	38±7	17.76	-0.64	0.73
Clays	10	66	29±8	27.54	0.37	0.10
pH	4.47	8.14	7.00±0.93	13.35	-0.97	-0.34
OM	0.90	132.30	48.00±14.38	29.95	0.56	1.14
TN	1.06	5.26	2.66±0.71	26.76	0.41	0.10
AN	64.70	447.40	202.98±54.07	26.64	0.46	0.51
TP	0.27	2.84	0.92±0.28	30.28	0.32	1.78
AP	1.66	118.80	36.48±17.74	48.62	0.89	1.40
TK	18.60	725.70	205.71±87.51	42.54	0.68	1.25
AK	6.22	40.10	12.74±3.83	30.05	2.33	10.16
Ca <sup>2+</sup>	2.11	83.77	33.27±23.35	70.19	0.62	-1.00
Mg <sup>2+</sup>	0.08	7.55	1.65±1.04	63.21	1.25	2.02
S	7.40	594.57	39.42±35.83	90.89	8.26	107.76
Cl <sup>-</sup>	0.00	98.09	6.30±9.84	156.19	2.72	11.96
B	0.19	1.36	0.55±0.18	32.59	1.05	1.88
Fe	10.80	502.10	142.86±89.59	62.71	1.23	1.23
Mn	0.81	294.20	33.24±31.53	94.87	2.94	12.37
Cu	0.27	96.30	4.70±5.31	113.05	12.66	190.18
Zn	0.42	233.00	4.44±9.77	220.12	15.65	318.23
Mo	0.00	4.40	0.16±0.23	143.79	8.18	124.05

Note: Coarse sands, Fine sands, Silts and Clays, %; OM, TN, TP and TK, g kg<sup>-1</sup>; AN, AP, AK, S, Cl<sup>-</sup>, B, Fe, Mn, Cu, Zn and Mo, mg kg<sup>-1</sup>; Ca<sup>2+</sup>, coml(1/2Ca<sup>2+</sup>) kg<sup>-1</sup>; Mg<sup>2+</sup> coml(1/2Mg<sup>2+</sup>) kg<sup>-1</sup>.

**Table 5.** Pearson correlation coefficients between CEC and other properties.

Soil property	CS	FS	Silt	Clay	pH	OM	TN	AN	TP	AP	TK
Total 1055	R	-.358**	-.445**	.277**	.437**	.327**	.041	.049	-.010	.312**	.106**
	S	.000	.000	.000	.000	.000	.188	.111	.753	.000	.001
GY 560	R	-.330**	-.375**	.046	.477**	.076	-.042	-.009	-.051	.102*	.028
	S	.000	.000	.280	.000	.074	.320	.836	.229	.016	.505
YX 115	R	-.605**	-.339**	.416**	.321**	.371**	.095	.154	.180	.384**	.219*
	S	.000	.000	.000	.000	.000	.312	.101	.054	.000	.019
JH 110	R	-.052	-.350**	.029	.236*	.151	-.040	-.045	-.174	.198*	.175
	S	.592	.000	.764	.013	.116	.677	.638	.069	.038	.068
SX 45	R	-.590**	-.398**	.591**	.513**	.430**	.231	.239	.259	-.361*	-.222
	S	.000	.007	.000	.000	.003	.127	.113	.086	.015	.143
AR 100	R	-.361**	-.412**	.514**	.524**	.603**	.317**	.358**	.175	.407**	-.266**
	S	.000	.000	.000	.000	.000	.001	.000	.081	.000	.007
YZ 96	R	-.446**	-.215*	.367**	.170	.067	.098	.139	.133	.326**	.201*
	S	.000	.035	.000	.098	.517	.343	.175	.197	.001	.049

**Table 5.** Pearson correlation coefficients between CEC and other properties (Continued)

Soil property		AK	Ca <sup>2+</sup>	Mg <sup>2+</sup>	S	Cl <sup>-</sup>	B	Fe	Mn	Cu	Zn	Mo
Total 1055	R	-.123**	.183**	.203**	-.026	-.134**	.136**	-.268**	.162**	.040	-.051	.071*
	S	.000	.000	.000	.395	.000	.000	.000	.000	.189	.096	.021
GY 560	R	-.004	-.027	.041	-.078	-.094*	.017	-.122**	.123**	-.043	-.087*	.047
	S	.925	.516	.332	.066	.026	.691	.004	.003	.308	.040	.270
YX 115	R	.081	.260**	.253**	-.042	-.176	.321**	-.213*	-.097	.335**	-.122	.068
	S	.392	.005	.006	.656	.060	.000	.022	.304	.000	.196	.469
JH 110	R	.182	.046	-.060	-.246**	-.139	.258**	-.141	-.111	.076	-.103	-.121
	S	.057	.633	.530	.010	.147	.007	.141	.250	.433	.283	.207
SX 45	R	-.072	.528**	.640**	.201	.277	.671**	-.295*	-.046	-.226	-.351*	-.217
	S	.641	.000	.000	.185	.066	.000	.049	.764	.136	.018	.152
AR 100	R	.140	.603**	.446**	.227*	.192	.231*	-.486**	.093	.253*	-.018	.072
	S	.165	.000	.000	.023	.055	.021	.000	.358	.011	.863	.479
YZ 96	R	-.029	.119	.098	-.097	-.050	.274**	.074	-.198	.281**	-.136	-.058
	S	.777	.249	.344	.347	.629	.007	.473	.054	.006	.188	.576

Note: in the first line, CS means coarse sands, FS means fine sands; In the first column, the number following the region means the number of soil samples; In the second column, R, Pearson coefficient; S, sig.(2-tailed).

**Table 6.** Optimal regression equation between CEC and other properties in Chenzhou

Region	Optimal regression model	R <sup>2</sup>	RSME	RMSE/S.D.
Total	CEC=-0.606-0.357CS-0.133FS+0.336Clay+2.627pH+0.864TP	0.407	5.94	0.77
Guizhou	CEC=20.048-0.420CS-0.134FS+0.341Clay	0.288	6.45	0.85
Yongxing	CEC=30.655-0.593CS-0.353FS-0.342Silt+2.367pH+2.863TP-0.608B+0.977Cu	0.489	6.02	0.74
Jiahe	CEC=33.250-0.395FS	0.123	5.94	0.94
Anren	CEC=-1.484-0.222CS+0.110Silt+0.257Clay+0.545pH-0.023OM+1.216TN-2.110TP-0.202Ca <sup>2+</sup> +1.070Mg <sup>2+</sup> +0.05Fe	0.602	3.42	0.66
Yizhang	CEC=14.153-0.348CS+0.153Silt+0.112TK	0.231	5.14	0.89
Suxian	CEC=-5.821-0.350CS+0.288Silt+0.240Clay+0.936pH-5.630TP+0.099Ca <sup>2+</sup> +2.352Mg <sup>2+</sup> +0.020Zn	0.795	2.46	0.57

Note: CS, coarse sands; FS, fine sands; CEC model was setup for Beihu and Linwu due to their less samples (< 20).

## 4. Discussion

### 4.1. High CEC in tobacco-planting fields in Chenzhou

CEC was high in tobacco-planting fields in Chenzhou, the mean value of CEC was 22.05 cmol(+) kg<sup>-1</sup>, higher than the very high grade of CEC ( $\geq 20$  cmol(+) kg<sup>-1</sup>). The high CEC could be attributed to the high values of fine particles, pH and OM of the

samples, because many studies have proved well that CEC usually is positively correlated with clays, pH and OM, while negatively correlated with sands [5-14, 46-53], and Zhang and Zhu (1993) found that the positive contribution of silts to soil CEC could not be ignored [54]. From Table 4 it could be seen that both the contents of silts and clays were high, which were ranged from 10%~56% and 10%~66% with a mean of 38% and 29%, respectively, in total constituted of 2/3 of the particle composition. The high content of silts and clays could be attributed to that about 75% of the soil samples in Chenzhou were derived from the clayey parent materials of limestone and Quaternary red clay [42-43, 55]. pH value was also high in Chenzhou, ranged from 4.47 to 8.14 with a mean of 7.00, 89.18% or 62.37% of the soil samples were  $\geq 5.5$  or  $\geq 7.0$  in pH. High pH in Chenzhou could be attributed to the samples came from the paddy-planting and to high contents of  $\text{Ca}^{2+}$ , ranged from 2.11~83.77 cmol( $\text{Ca}^{2+}$ )  $\text{kg}^{-1}$  with a mean of 33.27 cmol( $\text{Ca}^{2+}$ )  $\text{kg}^{-1}$ , 85.12% of the samples were higher in  $\text{Ca}^{2+}$  ( $\geq 10$  cmol( $\text{Ca}^{2+}$ )  $\text{kg}^{-1}$ ). High  $\text{Ca}^{2+}$  was mostly due to superphosphate fertilizer applied for tobacco-rice, fired soil used to improve soil quality [56-57], and possibly  $\text{Ca}^{2+}$  was dissolved out of the limestone for most tobacco-planting fields in Chenzhou are located in the limestone hill and mountainous area [32]. Meanwhile, OM was also high in Chenzhou, ranged from 0.90 g  $\text{kg}^{-1}$  to 132.30 g  $\text{kg}^{-1}$  with a mean of 48.00 g  $\text{kg}^{-1}$ , and 86.06% and 90.33% of the samples were  $\geq 30$  g  $\text{kg}^{-1}$  in OM (high grade of OM). High OM in tobacco-planting fields in Chenzhou was decided by tobacco-rice rotation, straw returning to the field and organic fertilizer application [58-60].

#### 4.2. Necessity for New Transfer function of CEC

Previous studies found that the application of the existing pedotransfer function models was usually limited in a new region due to the different backgrounds of study regions [61-62]. For example, OM is an indispensable parameter in most existing CEC models because OM is usually significantly correlated with CEC [5-14]. However, in this study no significant correlation was found between OM and CEC in most regions of Chenzhou ( $R=0.040\sim0.231$ ,  $\text{Sig.}=0.127\sim6.77$ ) except Anren ( $R=0.317^{**}$ ,  $\text{Sig.}=0.001$ ), thus OM was not involved in CEC models established in this study except CEC model in Anren, which not only showed the particularity of the influencing factors of soil CEC in tobacco-planting fields in Chenzhou, but also proved again the necessity for establishing a new model for a new study region. The disappearance of OM from the CEC predicting model could also be attributed to narrower variation but high value of OM, the C.V. of OM was 61.86% of that of CEC, meanwhile, high OM more easily covers up the interchange points of cations in the process of organic-inorganic recombination [2, 63-64], both may weaken the positive contribution of OM to CEC.

#### 4.3. Scale Effects of CEC Predicting Model

Most soil properties have the scale effects, and usually the larger the study area, the more influencing factors, the greater the variability of the study object, and then the lower the universality of the model established [61-62], but sometimes the real situation is not always the case [65]. In our study, it could be found that scale down (from whole Chenzhou City to six regions in Chenzhou City) caused the accuracy of CEC model increased in Yongxing, Anren and Suxian (mainly located in northeast and central of Chenzhou) while decreased in Guiyang, Jiahe and Yizhang (mainly located

in west, southwest and south of Chenzhou), and usually the accuracy was low when less properties were significantly correlated with CEC, for examples,  $\leq 3$  properties (coarse sands and clays in Guiyang, fine sands in Jiahe, coarse sands, silts and total K in Yizhang) were significantly correlated with CEC, so the accuracies of their CEC models ( $R^2=0.795^{**}$ , 0.602 $^{**}$  and 0.489 $^{**}$ , and RMSE/S.D=0.57, 0.66 and 0.74, respectively) were lower than those of Yongxing, Anren and Suxian ( $R^2=0.795^{**}$ , 0.602 $^{**}$  and 0.489 $^{**}$ , and RMSE/S.D=0.57, 0.66 and 0.74, respectively), where  $\geq 5$  properties were significantly correlated with CEC, which not only reflected the spatial complexity and differences between different regions in Chenzhou, but also proves further that it would be better to setup the optimal CEC models for different regions.

#### 4.4. Influences of Climate Parameters and Parent Materials on Soil CEC

Bai et al. [66] found that climate parameters and parent materials had great influences on soil CEC, soil CEC had significant negative correlation with mean annual temperature (MAT,  $P<0.01$ ) and quadratic function with mean annual precipitation (MAP,  $P<0.01$ ), and CEC of soils derived from glacial drifts were higher than those of soils from alluvial sediments and purple siltstone ( $P<0.05$ ).

In this study there is only one meteorological station in each region of Chenzhou, thus, it is unreliable to use the climate data of the 8 stations to extract the information of climate parameters of each typical field through the method of spatial interpolation. However, the relationship between mean soil CEC with MAP and MAT in the eight regions was analyzed (see Table 7), and the results showed that Pearson correlation coefficients of soil CEC with MAP and MAT were -0.290 and -0.260 with Sig.(2-tailed) of 0.486 and 0.534, respectively, which indicated no significant correlation between soil CEC and MAP and MAT, so in this study, the two climate parameters were not used in soil CEC predication models.

**Table 7.** Mean values of soil CEC, annual precipitation (MAP) and air temperature (MAT).

Region	CEC cmol(+) kg <sup>-1</sup>	MAP Mm	MAT °C
Guiyang	23.80	1385	17.2
Yongxing	24.11	1417	17.6
Jiahe	23.83	1409	18.3
Anren	15.31	1404	17.7
Yizhang	18.10	1453	18.7
Suxian	16.30	1487	18.2
Beihu	20.88	1504	18.4
Linwu	15.78	1422	17.9

The parent material could affect soil CEC mainly through particle size distribution and clay mineral composition of the derived soil [1, 4]. The 1055 typical fields in this study are all under tobacco and rice rotation, and all located in the flat terrains along the rivers and valleys, thus, their soil parent materials all are the alluvial materials, but the material sources of these alluvial materials might be different, which could be roughly divided into three types: limestone (limestone, dolomite, slate, etc., 883 typical fields), Quaternary red clays (41 typical fields) and sandstones (sandstone, conglomerate, granite, etc., 130 typical fields). Table 8 shows that soil CECs of

limestone and Quaternary red clay were significantly higher than that of sandstone ( $P<0.01$  and  $P<0.05$ , respectively), which could be attributed to that clay contents were significantly higher while sand contents were significantly lower in the former two than that of the latter (see Table 9), because as shown in Table 5 that soil CEC was significantly positively correlated with clay content but negatively correlated with sand contents (coarse and fine sands).

**Table 8.** CECs of soils with different parent materials.

Parent material	Sample no.	Minimum	Maximum	Mean±S.D.	Skewness	Kurtosis
Limestone	884	3.50	48.50	23.95±7.48 A	0.46	0.03
Quaternary red clay	41	7.10	37.20	20.90±7.31 a	0.06	-0.30
Sandstones	130	5.80	42.40	16.35±6.78 b	1.11	1.98

**Table 9.** Contents of clays (<0.002mm) and sands (2~0.02 mm) in soils with different parent materials.

	Parent material	Minimum	Maximum	Mean±S.D.	Skewness	Kurtosis
Clays	Limestone	9.64	66.05	28.86±8.06 a	0.37	0.09
	Quaternary red clay	19.48	46.98	32.21±6.88 A	0.23	-0.45
	Sandstone	13.50	44.29	26.48±6.44 b	0.21	-0.19
Sands	Limestone	12.57	78.56	32.08±8.35 C	12.57	78.56
	Quaternary red clay	19.00	61.45	32.16±8.33 b	19.00	61.45
	Sandstone	7.48	68.09	41.81±12.23 a	7.48	68.09

## 5. Conclusion

This study disclosed that soil CEC averagely was high in tobacco-planting fields in Chenzhou, more than half of the tobacco fields in Chenzhou were in the very high grade of soil CEC, there were significant differences in soil CEC among different regions in Chenzhou. Different soil properties were most important for CEC predicting models in different regions, and the optimal soil CEC predicting models were different in different regions, the sampling and study region must be considered in establishing or applying the optimal soil CEC models. Soil organic matter is not a variable in soil CEC predicting models for tobacco-planting fields in Chenzhou.

## Acknowledgements

This study was supported by the Project of Chenzhou Company of Hunan Tobacco Company (No. 2019-45). We would like to express thanks to those for soil sampling and analysis.

## References

- [1] Huang CY. Pedology. China Agriculture Press, Beijing, 2000.
- [2] Li XY. Soil Chemistry. High Education Press, Beijing, 2001.
- [3] Cooperative research group on Chinese soil taxonomy. Chinese Soil Taxonomy. Science Press, Beijing and New York, 2001.
- [4] Soil Survey Staff 2014a Keys to Soil Taxonomy, 12th Edition. USDA, Washington D C., 2014.
- [5] Raha NS, Alhumairi BAJ. Modelling of soil cation exchange capacity for some soils of east gharaf lands from mid-Mesopotamian plain (Wasit province/Iraq). *Int. J. Environ. Sci. Tech.* 2019; 16: 3183-3192.
- [6] Khaledian Y, Brevil EC, Pereira P, Cerdà A, Fattah MA, Tazikeh H. Modeling soil cation exchange capacity in multiple countries. *Catena* 2017; 158: 194-200.
- [7] Shiri J, Keshavarzi A, Kisi O, Iturraran-Viveros U, Bagherzadeh A, Mousavi R, Karimi S 2017 Modeling soil cation exchange capacity using soil parameters: Assessing the heuristic models. *Comput. Electron. Agr.* 2017; 135: 242-251.
- [8] Seyedmohammadi J, Esmaelnejad L, Ramezanpour H. Determination of a suitable model for prediction of soil cation exchange capacity. *Model. Earth Syst. Environ.* 2016; 2:156. DOI 10.1007/s40808-016-0217-4
- [9] Liao K, Xu S, Zhu Q. Development of ensemble pedotransfer functions for cation exchange capacity of soils of Qingdao in China. *Soil Use Manag.* 2015; 31: 483-490.
- [10] Obalum SE, Watanabe Y, Igwe CA, Obi ME, Wakatsuki T. Improving on the prediction of cation exchange capacity for highly weathered and structurally contrasting tropical soils from their fine-earth fractions. *Commun Soil Sci. Plan. Ana.* 2013; 44: 1831-1848.
- [11] Yukselen Y, Kaya A. Prediction of cation exchange capacity from soil index properties. *Clay Miner.* 2006; 41(4): 827-837.
- [12] Seybold CA, Grossman RB, Reinsch TG. Predicting cation exchange capacity for soil survey using linear models. *Soil Sci. Soc. Am. J.* 2005; 69, 856-863.
- [13] Krogh LH, Breuning M, Greve HM. Cation-exchange capacity pedotransfer functions for Danish soils. *Acta Agri. Scandinavica Sec. B - Soil Plant Sci* 2002; 50(1): 1-12.
- [14] Manrique LA, Jones CA, Dyke PT. Predicting cation-exchange capacity from soil physical and chemical properties. *Soil Sci. Soc. Am. J.* 1991; 55: 787-794.
- [15] Tang LN, Chen SH, Li ZB, Zeng WL, Liu XG, Jiang ZM. Main nutrients in and fertilization solutions to soil of Fujian tobacco-growing area. *Tobac. Sci. Technol.* 2008; (1): 56-60.
- [16] Yin YQ, Luo BX, Wei ZY, Kong F. Analysis on soil fertility characteristics for tobacco planting in Hechi district of Guangxi. *Guangdong Agri. Sci.* 2012; (16): 84-87.
- [17] Ye XF, Ling AF, Li YJ, Yang YX, Huang YJ, Chen XH, Liu GS. Evaluation of the soil fertility in tobacco-growing areas in Henan Province. *Chin. J. Soil Sci.*, 2009; 40(6): 1303-1307.
- [18] Song DT, Yao XH, Guo QZ. Analysis of soil nutrients and fertilization technology in Liangshan tobacco growing area. *Acta Agri. Jiangxi*, 2012; 24(12): 108-110.
- [19] Li SL, Chen JB, Wu WB, Kee YS, Zhan ZS, Yuan QH, Peng WS, Tan MX, Zhang ZC, Liu Y. Soil analysis to major soil in Nanxiong tobacco-growing regions. *Guangdong Agri. Sci.*, 2011; Suppl: 103-107.
- [20] Ma Y, Chen XH, Li HX. Analysis of chemical properties and nutrient status of tobacco growing soil in Qianxinan Prefecture. *Till. Culti.*, 2009; (1): 29-30.
- [21] Yuan F, Jiang DZ, Zhang YY. Study on nutrition of tobacco-growing soil in Yongzhou. *Hunan Agri. Sci.* 2009; (4): 55-57, 61.
- [22] Song ZX, Gao M, Guan BQ, Xu AD. Soil fertility properties in the tobacco-planted area of Chongqing. *Chin. J. Soil Sci.* 2005; 36(5): 664-668.
- [23] Qin ZL. Study on the fertility characteristics of planted tobacco soil in Guizhou. Master's Thesis, Southwest Uni., Chongqing, 2007.
- [24] Li Q, Yan CB, Liu YJ, Li J, Rang ZW, Xiao YS, Li HG, Peng SG. Preliminary study on spatial distribution and influencing factors of tobacco-growing soil pH in Chenzhou. *Acta Tabac. Sin.* 2019; 25(3). doi: 10.16472/j.chinatobacco.2018.276.
- [25] Chen YL. Study on the soil acidity characteristics and influencing factors of tobacco-planting areas in Sichuan Province. Master's Thesis, Sichuan Agri. Uni., Chengdu, 2015.
- [26] Liu RT, Zhao SY, Huang YL. Potash status of major tobacco soils in Hunan Province, *J. Hunan Agri. Col.* 1991; 17(Suppl): 312-318.
- [27] Hai HH, Zhu JF, Huang HT, Xu ZC. Relationships of soil available potassium content with soil factors and quality traits of tobacco in middle Henan Province. *Acta Agri. Jiangxi* 2015; 27(2): 54-57.
- [28] Chang NJ, Liu QL, Li ZH, Li JY, Ma ED, Zhang YG. Grey correlation degree between ecological factors and quality of flue-cured tobacco in typical tobacco growing areas. *Southwest Chin. J. Agri. Sci.* 2017; 30(8): 1754-1759.

- [29] Wang H. Liangshan tobacco-growing soil nutrient impact on the quality of tobacco leaf. Master's Thesis, Henan Agri. Uni., Zhengzhou, 2012.
- [30] Li YY, Qin GR, Wang L, Yu J, Xu RB, Peng WX, Rao XF. Soil nutrient elements affecting the occurrence of tobacco bacterial wilt in Qingjiang River Basin. *J. Southern Agri.* 2018; 49(4): 656-661.
- [31] Gao FH, Chen J, Tao Q, Xu Y, Li ZH, Yang SS, Peng Y, Ju FQ. Effects of physicochemical properties of tobacco planting soil on occurrence of tobacco diseases. *Hunan Agri. Sci.* 2012; (9): 74-76.
- [32] Editorial Committee of Tobacco Annals of Chenzhou City 2005 *Tobacco Annals of Chenzhou City*. Hunan People's Publishing House, Changsha, 2005.
- [33] Luo DS, Wang B, Qiao XY. Explanation of national regionalization of leaves style of flue-cured tobacco. *Acta Tabac. Sin.* 2019; 25(4): 1-9.
- [34] Gao X, Zhou LK, Guo T, Li HG, Deng XH, Xiao YS, Li Q. Spatial and temporal variability of soil organic matter and total nitrogen and influencing factors in Chenzhou tobacco-growing area, Hunan Province. *Chin. J. Soil Sci.* 2020; 51(3): 686-693.
- [35] Guo T, Li HG, Li W, Peng SG, Xiao CS, Li J, Liu YJ, Xiao YS, Li LJ, Li Q. Spatial-temporal variability and factors to influence boron availability in tobacco-planting soils of rotating paddy-upland fields in southern Hunan. *Tobac. Sci. Technol.* 2019; 52(7): 27-34.
- [36] Li Q, Yan CB, Liu YJ, Li J, Rang ZW, Xiao YS, Li HG, Peng SG. Preliminary study on spatial distribution and influencing factors of tobacco-growing soil pH in Chenzhou. *Acta Tabac. Sin.* 2019; 25(3). doi: 10.16472/j.chinatobacco.2018.276.
- [37] Kuang CF, Zhou GS, Deng ZP, Li XY, Cheng JP, Shi XB, Xiao HQ; Li MD. Soil nutrient status in Chenzhou tobacco planting areas. *Acta Tabac. Sin.* 2020; 31(3): 33-37.
- [38] Zhang GL, Gong ZT. *Soil Survey Laboratory Methods*. Science Press, Beijing, 2012.
- [39] Bao SD. *Soil Agrochemical Analysis* (the 3rd edition). China Agriculture Press, Beijing, 2000.
- [40] Lu RK. *Methods for agricultural chemical analysis of soils*. China Agricultural Science and Technology Press, Beijing, 2000.
- [41] Agricultural Chemistry Specialized Committee of Chinese Society of Soil Science 1983 *Methods for Conventional Analysis of Soil Agrochemistry*. Science Press, Beijing. 1983.
- [42] Li QK, Xiong Y. *Soils of China*. 2nd Edition, Science Press, Beijing, 1990.
- [43] Soil survey office of China. *Soils of China*. China Agriculture Press, Beijing, 1998.
- [44] Yu JH, He XH. *Data statistical analysis and SPSS applications*. Posts & Telecommunications Press, Beijing, 2003.
- [45] Nielsen DR, Bouma J. Soil spatial variability. Proceedings of a workshop of the ISSS and SSSA, Las Vegas, USA/Pdc296. Pudoc Netherlands: Wageningen, 1985.
- [46] Bai ZQ, Zhang SR, Zhong QM, Wang GY, Xu GR, Ma XJ. Characteristics and impact factors of soil cation exchange capacity (CEC) in western margin of Sichuan Basin. *Soils* 2020, 52(3): 581-687.
- [47] Liao K, Xu S, Zhu Q. Development of ensemble pedotransfer functions for cation exchange capacity of soils of Qingdao in China. *Soil Use Manage.* 2015; 31: 483-490.
- [48] Wang WY, Zhang LP, Liu Q. Distribution and affecting factors of soil cation exchange capacity in watershed of the Loess plateau. *J. Soil Water Conserv.* 2012; 26(5): 123-127.
- [49] Cheng XF, Zhu H, Hao LX, Shi XZ. Spatial predication of soil cation exchange capacity (CEC) in hilly region. *Chin. J. Appl. Environ. Biol.* 2008; 14(4): 484-487.
- [50] Liu SQ, Pu YL, Zhang SR, Wang CQ, Deng LJ. Spatial change and affecting factors of soil cation exchange capacity in Tibet. *J. Soil Water Conserv.* 2004; 18(5): 1-5.
- [51] Zhao ZZ. Study on relationship between organic matter, soil fractions and CEC in Qinghai soil. *Sci. Technol. Qinghai Agri. Fores.* 2004; (4): 4-6.
- [52] Xu MG, Zhang JX, Zhang H, An ZS. Study on the influencing factors of cation exchange capacity of Heilutu and yellow-brown soil. *Chin. J. Soil. Sci.* 1991; 22(3): 108-110.
- [53] Xu MG, An ZS. A study on the relationship between the mineral compositions and the cation exchange capacity in Shaanxi soils. *Acta Uni. Sep. Occident. Agri.* 1988; 2(6): 310 -313.
- [54] Zhang MK, Zhu Z X. Effect of silts on cation exchange capacity in soil. *Soil Fertil.* 1993; (4): 41-43.
- [55] Xu JP, Li JG, Wei YJ, Zhang GH, Yang BG, Cai CF. Fractal characteristics of particle composition for soils developed from different parent materials. *Acta Pedo. Sin.* 2020; 57(5): 1197-1205.
- [56] Li XC, Li HG, Xiao YS, Song WJ, Zheng XB, Dong JX. Effects of fired soil on root growth and rhizosphere nutrients of tobacco at seedling stage. *Chin. Tobac. Sci.* 2020; 41(5): 43-48.
- [57] Kuang CF, Li HG, Xu QX, Li XY, Huang CY, Fang M, Luo JR, Li YX. Study on improving rhizosphere environment of flue-cured tobacco by applying ash soil. *Agri. Devel. Equip.* 2013; (9): 52-53.
- [58] Gao X, Zhou LK, Guo T, Li HG, Deng XH, Xiao YS, Li Q. Spatial and temporal variability of soil organic matter and total nitrogen and influencing factors in Chenzhou tobacco-growing area, Hunan Province. *Chin. J. Soil Sci.* 2020; 51(3): 686-693.

- [59] Wang XD, Zhuang JJ, Liu BY, Li SS, Zhao X, Liu Y, Zhang HL. Residue returning induced changes in soil organic carbon and the influential factors in China's croplands: A meta-analysis. *J. Chin. Agri. Uni.* 2020; 25(8)L 12-24.
- [60] Jin ZL, Jia SC, Peng SG, Liu YJ. Effects of the application level of cake fertilizer on the growth and quality of flue cured tobacco in high organic matter-field of south Hunan. *Crop Res.* 2018; 32(6): 516-520.
- [61] Van LK, Bouma J, Herbst M, Koestel J, Minasny B, Mishra U, Montzka C, Nemes A, Pachepsky YA, Padarian J, Schaap MG, Tóth B, Verhoef A, Vanderborght J, Van der Ploeg MJ, Weihermüller L, Zacharias S, Zhang YG, Vereecken H. Pedotransfer functions in Earth system science: Challenges and perspectives. *Rev. Geophys.* 2017; 55: 1199–1256.
- [62] McBratney AB, Minasny B, Cattle SR, Vervoort RW. From pedotransfer functions to soil inference systems. *Geoderma* 2002; 109: 41– 73
- [63] Desilva JA, Toth ST. Cation-exchange reactions, electrokinetic and viscometric behavior of clay-organic complexes. *Soil Sci.* 1964; 97(1): 63-73.
- [64] Ji GL. Soil electrochemical properties and their research methods (revised edition). Science Press, Beijing, 1980.
- [65] Ge C, Liu HL, Nie CJ, Shen Q, Zhang SW. Scale effect of soil fertility spatial variability and its influencing factors. *Resour. Sci.* 2019; 41(4): 753-765.
- [66] Bai ZQ, Zhang SR, Zhong QM, Wang GY, Xu GR, Ma XJ. Characteristics and impact factors of soil cation exchange capacity (CEC) in western margin of Sichuan basin. *Soils* 2020; 52(3): 581-587.

# Cutting L1-Norm Distance Discriminant Analysis with Sample Reconstruction

Guowan SHAO<sup>a,1</sup>, Chunjiang PENG<sup>a</sup>, Wenchu OU<sup>a</sup> and Kai DUAN<sup>b</sup>

<sup>a</sup>College of Mechanical and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411201, Hunan, PR China

<sup>b</sup>Engineering Research Center for Advanced Mine Equipment of Ministry of Education, Hunan University of Science and Technology, Xiangtan 411201, Hunan, PR China

**Abstract.** Dimensionality reduction plays an important role in the fields of pattern recognition and computer vision. Recursive discriminative subspace learning with an L1-norm distance constraint (RDSL) is proposed to robustly extract features from contaminated data and L1-norm and slack variables are utilized for accomplishing the goal. However, its performance may decline when too many outliers are available. Moreover, the method ignores the global structure of the data. In this paper, we propose cutting L1-norm distance discriminant analysis with sample reconstruction (C-L1-DDA) to solve the two problems. We apply cutting L1-norm to measure within-class and between-class distances and thus outliers may be strongly suppressed. Moreover, we use cutting squared L2-norm to measure reconstruction errors. In this way, outliers may be constrained and the global structure of data may be approximately preserved. Finally, we give an alternating iterative algorithm to extract feature vectors. Experimental results on two publicly available real databases verify the feasibility and effectiveness of the proposed method.

**Keywords.** Discriminant analysis, dimensionality reduction, cutting L1-norm, sample reconstruction

## 1. Introduction

Dimensionality reduction is utilized for preprocessing high dimensional data, which plays an irreplaceable role in the fields of pattern recognition and computer vision. In the past decades, it has been extensively studied and a variety of methods have been presented, such as linear discriminant analysis (LDA) [1], locality preserving projections [2] and principal component analysis (PCA) [3]. Among them, LDA and PCA are two basic methods, which are widely used in practice. The former is a supervised method and based on the criterion that the distances between samples from the same class are minimized and the distances between samples from different classes are maximized simultaneously.

LDA is an effective method for dimensionality reduction. However, its performance may decline when the data are contaminated by outliers or noise. Many methods have

---

<sup>1</sup> Corresponding Author, Guowan SHAO, College of Mechanical and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411201, Hunan, PR China; E-mail: gwshao\_ezhou@163.com, 821211624@qq.com.

been presented to deal with the problem and enhance the performance [4-6]. From the view of mathematics, L1-norm is more robust to outliers or noise than L2-norm and hence some methods replace the latter with the former and improve the robustness [7-11]. Rotational invariant LDA (RILDA) [7] constructs a unified rotational invariant framework with L2,1-norm and generalizes Laplacian eigenmaps for better robustness. Based on LDA, L1-norm LDA [8] substitutes L1-norm for L2-norm and outperforms RILDA. However, the gradient ascending iterative algorithm may not guarantee to obtain an optimal solution. L1-LDA [9] follows the theoretical framework of Bayes optimality, but the method actually cannot guarantee the Bayes optimality for it introduces the alternating optimization strategy to obtain a solution. Moreover, it is not more robust to outliers or noise for the same reason. In order to avoid the problems, non-greedy L1-LDA [10] uses a non-greedy iterative algorithm to optimize the trace ratio form of L1-LDA. However, the method is sensitive to initialization for the introduction of the learning rate, and it is true to L1-norm distance-based effective LDA [11].

Recently, a novel recursive discriminative subspace learning method with an L1-norm distance constraint (RDSL) has been proposed to obtain a robust and discriminative subspace [12]. The method formulates LDA with the maximum margin criterion and avoids the introduction of the learning rate. It introduces slack variables in order to obtain better robustness. However, its performance may decline when too many outliers are available in training samples. Moreover, the method ignores the global structure of the data and thus there is still room for performance improvement.

In this paper, we present cutting L1-norm distance discriminant analysis with sample reconstruction (C-L1-DDA) to solve the two problems existing in RDSL. Based on RDSL, we apply cutting L1-norm to measure within-class and between-class distances, thus strongly suppressing outliers. Furthermore, we construct a model of sample reconstruction as a regularization term, which uses cutting squared L2-norm to measure reconstruction errors. The model can also suppress outliers and approximately preserve the global structure of data. Finally, we use an alternating iterative technique to optimize the objective function. Experimental results on two publicly available real datasets demonstrate that our method obtains better classification performance than RDSL.

## 2. Recursive Discriminative Subspace Learning

As previously mentioned, RDSL reformulates LDA according to the maximum margin criterion and avoids the sensitiveness to initialization. Here we give a brief review of RDSL for our method is based on it.

Let there be a data set  $X \in R^{m \times n}$ , which includes  $n$  samples from  $c$  classes in an  $m$  dimensional space. Let  $\bar{x}_i$  be the mean of class  $i$ ,  $i=1,2,\dots,c$ , and  $\bar{x}$  the mean of all the classes. Let  $x_{ij}$  be the  $j$  th sample from class  $i$  and  $n_i$  the number of samples from the same class. Then RDSL is formulated as

$$\begin{aligned} & \min_{w_q, \xi_i} \sum_{i=1}^c \sum_{j=1}^{n_i} \left\| w_q^T (x_{ij} - \bar{x}_i) \right\|_1 + \lambda \sum_{i=1}^c \xi_i \\ & \text{s.t. } \left\| w_q^T (\bar{x}_i - \bar{x}) \right\|_1 + \xi_i \geq 1 \\ & \quad \xi_i \geq 0, \quad i=1,2,\dots,c \end{aligned} \quad (1)$$

where  $w_q$  is the  $q$  th projection vector,  $1 \leq q \leq d$ , and  $d$  is the dimensionality of the subspace. Accordingly, the projection matrix  $W = [w_1, w_2, \dots, w_d]$ . Moreover,  $\xi_i$  is a slack variable, which is used to improve the generalization or the robustness of RDSL, and  $\lambda$  is the regularization parameter.

RDSL simultaneously considers within-class compactness and between-class separability and uses L1-norm and slack variables to improve the robustness to outliers or noise. However, it is difficult to directly solve Eq. (1) for the use of L1-norm. Fortunately, the problem can be transformed into a linear or quadratic programming form and then an iterative technique is used to obtain the solution [12].

### 3. Cutting L1-Norm Distance Discriminant Analysis

RDSL can obtain excellent performance when there are a few outliers in training samples. However, its performance may decline when too many outliers are available since, in this situation, L1-norm and slack variables may not eliminate the adverse effects of the outliers to a large extent.

Inspired by the success of the cutting L1-norm loss function in support vector machines [13], we first apply cutting L1-norm to RDSL in order to eliminate the effects of outliers. Our method is formulated as follows:

$$\begin{aligned} & \min_{w_q, \xi_i} \sum_{i=1}^c \sum_{j=1}^{n_i} \min \left( s, \|w_q^T (x_{ij} - \bar{x}_i)\|_1 \right) + \lambda \sum_{i=1}^c \xi_i \\ & \text{s.t. } \min \left( h, n_i \|w_q^T (\bar{x}_i - \bar{x})\|_1 \right) + \xi_i \geq 1 \\ & \quad \xi_i \geq 0, \quad i = 1, 2, \dots, c \end{aligned} \quad (2)$$

where  $s$  and  $h$  are two given values,  $s > 0$ ,  $h > 0$ .

By applying cutting L1-norm in the objective function and the constraint, our method can strongly suppress outliers. On the one hand, the distances between outliers and their class means are usually larger than those between normal samples and the means. The larger distances may dominate the objective function when too many outliers are available, which may lead to the drift of projection directions. It may be the reason that the performance of RDSL declines when training samples include too many outliers. We use cutting L1-norm to constrain the larger distances and thus may greatly improve the robustness to outliers. On the other hand, the distances between the class means and the mean of all the classes may also be severely influenced by outliers. In other words, the distances may be significantly increased. In this case, the drift of projection directions may occur for the constraint in Eq. (1). Therefore, we apply cutting L1-norm to constrain the terms with larger distances, which may alleviate the problem.

### 4. Cutting L1-Norm Distance Discriminant Analysis with Sample Reconstruction

RDSL is designed to robustly learn a series of projection vectors from noisy data. It realizes within-class compactness by the objective function and between-class separability by the constraint. However, the method ignores the global structure of the

data. As a result, the projection vectors may not extract as much discriminative information as possible.

In order to solve the problem, we utilize sample reconstruction for preserving the global structure. More specifically, we minimize the reconstruction errors of training samples, and cutting L1-norm is applied to accomplish the goal. The modified form is formulated in the following:

$$\begin{aligned} \min_{w_q, u_q, \xi_i} & \sum_{i=1}^c \sum_{j=1}^{n_i} \min \left( s, \|w_q^T (x_{ij} - \bar{x}_i)\|_1 \right) + \alpha \sum_{i=1}^c \sum_{j=1}^{n_i} \min \left( k, \|x_{ij} - u_q w_q^T x_{ij}\|_1 \right) + \lambda \sum_{i=1}^c \xi_i \\ \text{s.t. } & \min \left( h, n_i \|w_q^T (\bar{x}_i - \bar{x})\|_1 \right) + \xi_i \geq 1 \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, c \end{aligned} \quad (3)$$

where  $u_q$  is the  $q$ th reconstruction vector,  $1 \leq q \leq d$ ,  $\alpha$  is a regularization parameter,  $\alpha \geq 0$ , and  $k$  is a given value,  $k > 0$ . The reconstruction regularization term takes the form of cutting L1-norm and thus can improve the robustness to a certain extent. However, the form may lead to a high computational cost in the optimization when the number of training samples is large.

Therefore, we replace L1-norm with squared L2-norm in order to reduce the computational cost and the modified version is described as follows:

$$\begin{aligned} \min_{w_q, u_q, \xi_i} & \sum_{i=1}^c \sum_{j=1}^{n_i} \min \left( s, \|w_q^T (x_{ij} - \bar{x}_i)\|_1 \right) + \alpha \sum_{i=1}^c \sum_{j=1}^{n_i} \min \left( k, \|x_{ij} - u_q w_q^T x_{ij}\|_2^2 \right) + \lambda \sum_{i=1}^c \xi_i \\ \text{s.t. } & \min \left( h, n_i \|w_q^T (\bar{x}_i - \bar{x})\|_1 \right) + \xi_i \geq 1 \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, c \end{aligned} \quad (4)$$

Although L2-norm is sensitive to noise, it is possible to evade the problem in the form of cutting squared L2-norm. That is to say, the samples with larger reconstruction residuals are regarded as outliers and constrained.

## 5. Optimization

Eq. (4) involves the variables:  $w_q$ ,  $u_q$  and  $\xi_i$  and we adopt the alternating iterative method [14] to obtain the solutions.

### 5.1. Reformulation of Minimization Terms

Eq. (4) involves more complex minimization terms and it is difficult to directly optimize the problem. In order to solve the difficulty, we reformulate the terms in the following.

We first reformulate the constrain in Eq. (4). Let  $p_i = n_i (\bar{x}_i - \bar{x})$ , then

$$\begin{aligned} \min \left( h, n_i \|w_q^T (\bar{x}_i - \bar{x})\|_1 \right) &= \min \left( h, \|w_q^T p_i\|_1 \right) \\ &= \min \left( h, \max \left( 0, w_q^T p_i \right) \right) + \min \left( h, \max \left( 0, -w_q^T p_i \right) \right) \end{aligned}$$

and

$$\begin{aligned}\min(h, \max(0, w_q^T p_i)) &= \max(0, w_q^T p_i) - \max(0, w_q^T p_i - h) \\ \min(h, \max(0, -w_q^T p_i)) &= \max(0, -w_q^T p_i) - \max(0, -w_q^T p_i - h)\end{aligned}$$

Therefore, we have

$$\min(h, \|w_q^T p_i\|_1) = \|w_q^T p_i\|_1 - \max(0, w_q^T p_i - h) - \max(0, -w_q^T p_i - h)$$

Namely, the constraint term  $\min(h, n_i \|w_q^T(\bar{x}_i - \bar{x})\|_1) + \xi_i \geq 1$  can be transformed into as follows:

$$\|w_q^T p_i\|_1 - \max(0, w_q^T p_i - h) - \max(0, -w_q^T p_i - h) + \xi_i \geq 1$$

Let  $\pi_{i1} = \max(0, w_q^T p_i - h)$  and  $\pi_{i2} = \max(0, -w_q^T p_i - h)$ , then the constraint term can be written as follows:

$$\begin{aligned}\|w_q^T p_i\|_1 - \pi_{i1} - \pi_{i2} + \xi_i &\geq 1 \\ h + \pi_{i1} &\geq w_q^T p_i \\ h + \pi_{i2} &\geq -w_q^T p_i \\ \pi_{i1} &\geq 0, \pi_{i2} \geq 0\end{aligned}$$

Clearly,  $\|w_q^T p_i\|_1 = \text{sign}(w_q^T p_i) w_q^T p_i = \text{sign}(w_q^T p_i) p_i^T w_q$ , where  $\text{sign}(\cdot)$  is the sign function. Let  $a_i = \text{sign}(w_q^T p_i) p_i$ ,  $A_q = [a_1, a_2, \dots, a_c]^T$  and  $B = [p_1, p_2, \dots, p_c]^T$ . Let  $e_c$  denote a column vector and its elements consist of  $c$  ones. In addition, let  $\gamma_1 = [\pi_{11}, \pi_{21}, \dots, \pi_{c1}]^T$ ,  $\gamma_2 = [\pi_{12}, \pi_{22}, \dots, \pi_{c2}]^T$  and  $\gamma_3 = [\xi_1, \xi_2, \dots, \xi_c]^T$ . Accordingly, the constraint in Eq. (4) can be reformulated as the following form:

$$\begin{aligned}A_q w_q - \gamma_1 - \gamma_2 + \gamma_3 &\geq e_c \\ h e_c + \gamma_1 &\geq B w_q \\ h e_c + \gamma_2 &\geq -B w_q \\ \gamma_1 &\geq 0, \gamma_2 \geq 0, \gamma_3 \geq 0\end{aligned}$$

We then reformulate the first term of the objective function in Eq. (4). Let  $f_{ij} = x_{ij} - \bar{x}_i$ , then

$$\|w_q^T(x_{ij} - \bar{x}_i)\|_1 = \|w_q^T f_{ij}\|_1 = \frac{w_q^T f_{ij} f_{ij}^T w_q}{\|w_q^T f_{ij}\|_1}$$

Accordingly,

$$\min(s, \|w_q^T(x_{ij} - \bar{x}_i)\|_1) = \min(s, \|w_q^T f_{ij}\|_1) = \rho_{ij} w_q^T f_{ij} f_{ij}^T w_q + \theta_{ij} s$$

where  $\rho_{ij} = 1/\|w_q^T f_{ij}\|_1$  and  $\theta_{ij} = 0$  if  $\|w_q^T f_{ij}\|_1 < s$ , otherwise  $\rho_{ij} = 0$ ,  $\theta_{ij} = 1$ . Therefore, we have

$$\sum_{i=1}^c \sum_{j=1}^{n_i} \min(s, \|w_q^T(x_{ij} - \bar{x}_i)\|_1) = w_q^T F_q w_q + z_q$$

where  $F_q = \sum_{i=1}^c \sum_{j=1}^{n_i} \rho_{ij} f_{ij} f_{ij}^T$ , and  $z_q = \sum_{i=1}^c \sum_{j=1}^{n_i} \theta_{ij} s$ .

In addition, we reformulate the second term of the objective function. Clearly,

$$\left\| x_{ij} - u_q w_q^T x_{ij} \right\|_2^2 = x_{ij}^T x_{ij} - 2x_{ij}^T u_q x_{ij}^T w_q + w_q^T x_{ij} u_q^T u_q x_{ij}^T w_q$$

and

$$\min \left( k, \left\| x_{ij} - u_q w_q^T x_{ij} \right\|_2^2 \right) = (1 - \mu_{ij}) k + \mu_{ij} \left\| x_{ij} - u_q w_q^T x_{ij} \right\|_2^2$$

where  $\mu_{ij} = 1$  if  $\left\| x_{ij} - u_q w_q^T x_{ij} \right\|_2^2 < k$ , otherwise  $\mu_{ij} = 0$ . Accordingly,

$$\sum_{i=1}^c \sum_{j=1}^{n_i} \min \left( k, \left\| x_{ij} - u_q w_q^T x_{ij} \right\|_2^2 \right) = b_{q1} + a_{q1}^T w_q + w_q^T G_q w_q$$

where

$$\begin{aligned} b_{q1} &= \sum_{i=1}^c \sum_{j=1}^{n_i} ((1 - \mu_{ij}) k + \mu_{ij} x_{ij}^T x_{ij}) \\ a_{q1} &= -2 \sum_{i=1}^c \sum_{j=1}^{n_i} \mu_{ij} (x_{ij}^T u_q) x_{ij} \\ G_q &= \sum_{i=1}^c \sum_{j=1}^{n_i} \mu_{ij} x_{ij} u_q^T u_q x_{ij}^T \end{aligned}$$

## 5.2. Algorithm

After reformulating the constraint and the two terms of the objective function, we adopt an alternating iterative algorithm [14] to obtain the projection vector. Let  $w_q^t$  and  $u_q^t$  be the values of the projection vector and the reconstruction one at the  $t$ th iteration, and  $w_q^{(t+1)}$  and  $u_q^{(t+1)}$  the values at the  $(t+1)$ th iteration.

The iterative method alternately performs the following two main steps:

Step 1. Fix  $u_q$  and compute the value of  $w_q$  in Eq. (4). More specifically, we can obtain  $w_q^{(t+1)}$  by solving the following problem:

$$\begin{aligned} \min_{w_q^{(t+1)}, \gamma_1, \gamma_2, \gamma_3} & \left( w_q^{(t+1)} \right)^T (F_q + \alpha G_q) w_q^{(t+1)} + \alpha a_{q1}^T w_q^{(t+1)} + \lambda e_c^T \gamma_3 \\ \text{s.t. } & A_q w_q^{(t+1)} - \gamma_1 - \gamma_2 + \gamma_3 \geq e_c \\ & h e_c + \gamma_1 \geq B w_q^{(t+1)} \\ & h e_c + \gamma_2 \geq -B w_q^{(t+1)} \\ & \gamma_1 \geq 0, \gamma_2 \geq 0, \gamma_3 \geq 0 \end{aligned} \tag{5}$$

Eq. (5) is a quadratic programming problem and thus  $w_q^{(t+1)}$  can be easily obtained.

Step 2. Fix  $w_q$  and  $\xi_i$  and compute the value of  $u_q$  in Eq. (4). In this situation, Eq. (4) reduces to the following form:

$$\min_{u_q} \sum_{i=1}^c \sum_{j=1}^{n_i} \min \left( k, \left\| x_{ij} - u_q w_q^T x_{ij} \right\|_2^2 \right) \quad (6)$$

Accordingly,  $u_q^{(t+1)}$  can be easily obtained by solving the following quadratic programming problem:

$$\min_{u_q^{(t+1)}} g_q \left( u_q^{(t+1)} \right)^T u_q^{(t+1)} + a_{q2}^T u_q^{(t+1)} + b_{q2} \quad (7)$$

where

$$\begin{aligned} b_{q2} &= \sum_{i=1}^c \sum_{j=1}^{n_i} \left( (1 - \mu_{ij})k + \mu_{ij} x_{ij}^T x_{ij} \right) \\ a_{q2} &= -2 \sum_{i=1}^c \sum_{j=1}^{n_i} \mu_{ij} \left( x_{ij}^T w_q^{(t+1)} \right) x_{ij} \\ g_q &= \sum_{i=1}^c \sum_{j=1}^{n_i} \mu_{ij} \left( x_{ij}^T w_q^{(t+1)} \right)^2 \end{aligned}$$

After  $w_q$  is obtained by iteration, the training samples are updated by the factor  $(I_m - WW^T)$ , where  $I_m$  is the  $m \times m$  dimensional identity matrix [12]. The detailed process of the algorithm is shown in Table 1.

**Table 1.** Algorithm 1. The alternating iterative algorithm

<b>Input:</b> $X$ --- labeled data set, $X \in R^{m \times n}$ ; $\alpha, \lambda$ --- regularization parameters; $d$ --- number of projection vectors; <b>Output:</b> $W$ --- projection matrix, $W \in R^{md}$ .
<b>Procedure:</b> Initiate the projection matrix $W = \emptyset$ , where $\emptyset$ is the empty set; <b>For</b> $q = 1$ <b>to</b> $d$ <b>do</b> Initiate $w_q$ ; Compute $u_q$ according to Eq. (7); <b>While</b> not converge <b>do</b> Update $w_q$ according to Eq. (5); Update $u_q$ according to Eq. (7); <b>End while</b> Let $W = [W, w_q]$ , and normalize and orthogonalize it; Let $X = (I_m - WW^T)X$ ; <b>End for</b>

## 6. Experiments

We conduct experiments on USPS [15] and COIL-20 [16] datasets to evaluate LDA, coined complete large margin linear discriminant analysis (CLMLDA) [17], RILDA, RDSL and C-L1-DDA. The grayscale of images from the datasets is normalized to  $[0,1]$ .

After training samples are artificially contaminated and represented with the corresponding vectors, PCA is applied to reduce the dimensionality to  $n - c$ . In the experiments, the regularization parameters of CLMLDA, RDSL and C-L1-DDA are chosen from the set  $[10^{-3}, 10^{-2}, \dots, 10^3]$ . Our method has three cutting parameters, i.e.,  $h$ ,  $s$  and  $k$ , and it is difficult to specify them. As an alternative, the three parameters are empirically set in this way that five percent of training samples with the largest within-class distances are eliminated, which are regarded as outliers. The initial value of  $w_q$  is set to the solution of LDA. Basically, the convergence condition is set according to [12] and the difference lies in that the maximum number of iterations is set to fifty. In the experiments, the nearest neighbor classifier is employed in measuring classification performance, and it is directly used for preprocessed data by PCA as the baseline method. All the experiments are implemented ten times and the average lowest classification error rates are reported.

### 6.1. Experiments on the USPS Database

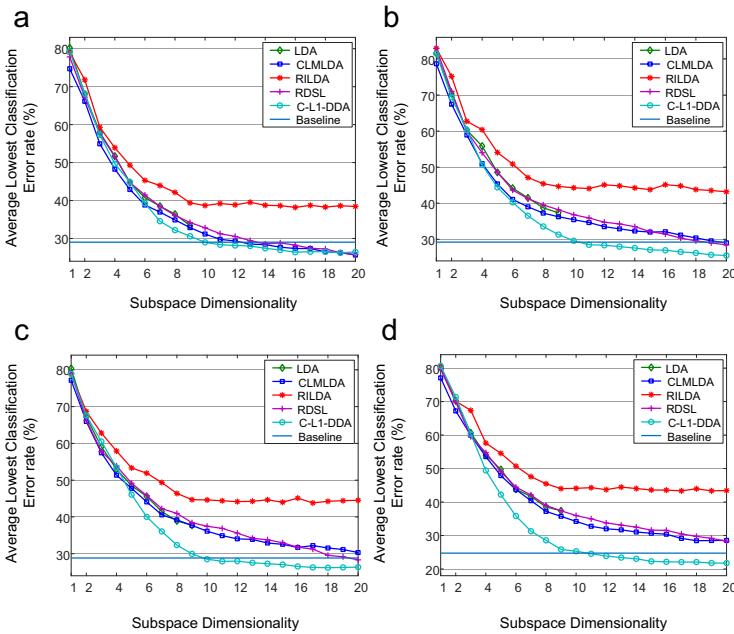
In this section, experiments are conducted with the USPS database. Ten handwritten digits from 0 to 9 of this database have 11,000 grayscale images whose size is  $16 \times 16$  pixels. A subset is first constructed for the experiments and each digit includes 100 images which are randomly selected from the original database. For each digit,  $K$  samples are selected for training and the remaining ones are used for testing. In the experiments,  $K$  is specified for  $\{10, 12, 14, 16\}$ . Forty percent of training samples are contaminated by inserting occlusion with black or white rectangular noise at a random location and the size of the rectangle is at least  $4 \times 4$  and up to  $10 \times 10$  pixels.

Average lowest classification error rates with different dimensions are shown in Figure 1. From the experimental results, we can conclude that the proposed method is significantly effective for classification.

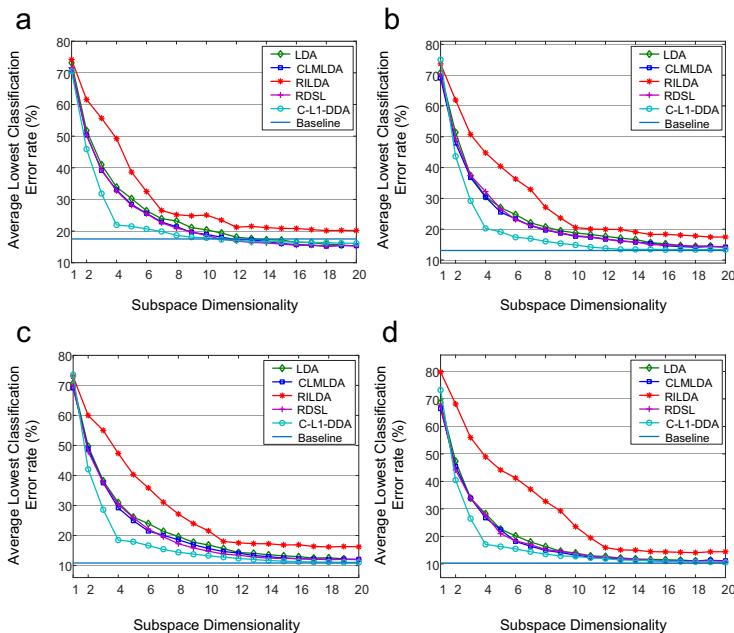
### 6.2. Experiments on the Coil-20 Database

The section deals with the experiments carried out on the Coil-20 database to test the performance of C-L1-DDA. The database consists of 1,440 grayscale images of twenty objects and each object has 72 images at an interval of five degrees of rotation. Each image is downsampled and resized to  $32 \times 32$  pixels. Let  $K = \{8, 10, 12, 14\}$ , and for each object,  $K$  samples are used for training and the rest for testing. Similar to the experiments on the USPS database, forty percent of training samples are contaminated and the difference is that the size of the noise rectangle is at least  $20 \times 20$  and up to  $30 \times 30$  pixels.

Average lowest classification error rates with different dimensions are given in Figure 2. From the figure we conclude that the performance of C-L1-DDA is better than that of RDSL



**Figure 1.** Average lowest classification error rates on the USPS database. (a) 10 samples, (b) 12 samples, (c) 14 samples, and (d) 16 samples.



**Figure 2.** Average lowest classification error rates on the Coil-20 database. (a) 8 samples, (b) 10 samples, (c) 12 samples, and (d) 14 samples.

## 7. Conclusions

In this paper, we present a novel discriminant analysis method, cutting L1-norm distance discriminant analysis with sample reconstruction (C-L1-DDA), which utilizes cutting L1-norm for measuring within-class distances and between-class distances. It may strongly suppress outliers and thus the robustness may be improved. Moreover, our method uses cutting squared L2-norm to measure reconstruction errors. As a result, it may suppress the outliers and approximately preserve the global structure of data. Finally, we use an alternating iterative algorithm to obtain projection vectors. Experimental results on two real databases demonstrate the effectiveness of C-L1-DDA. In the future, we will test the performance of our method on more databases.

## Acknowledgement

The work is supported by Natural Science Foundation of Hunan Province (Grant No.2018JJ2133).

## References

- [1] Belhumeur PN, Hespanha J, Kriegman J. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Trans Pattern Anal Mach Intell.* 1997 Jul; 19(7):711-20.
- [2] He X, Niyogi P. Locality preserving projections. In: Thrun S, Saul LK, Schölkopf B, editors. *Advances in neural information processing systems*; 2003 Dec 8-13; Cambridge, MA: MIT Press; c2004. p. 153-60.
- [3] Turk M, Pentland A. Eigenfaces for recognition. *Cognitive Neurosci.* 1991 Mar; 3(7):71-86.
- [4] Randles RH, Broffitt JD, Ramberg JS, Hogg RV. Generalized linear and quadratic discriminant functions using robust estimates. *J Amer Stat Assoc.* 1978 Sept;73(363):564-8.
- [5] Lanckriet GRG, El Ghaoui L, Bhattacharyya C, Jordan MI. A robust minimax approach to classification. *J Mach Learn Res.* 2002 Dec;3: 555-82.
- [6] Kim SJ, Magnani A, Boyd S. Robust Fisher discriminant analysis. In: Weiss Y, Schölkopf B, Platt J, editors. *Advances in Neural Information Processing Systems*; 2005 Dec 5-10; Cambridge, MA: MIT Press; c2006. p. 659-66.
- [7] Lai Z, Xu Y, Yang J, Shen L, Zhang D. Rotational invariant dimensionality reduction algorithms. *IEEE Trans Cybern.* 2017 Nov;47(11):3733-46.
- [8] Zhong F, Zhang J. Linear discriminant analysis based on L1-norm maximization. *IEEE Trans Image Process.* 2013 Aug;22(8):3018-27.
- [9] Zheng W, Lin Z, Wang H. L1-norm kernel discriminant analysis via Bayes error bound optimization for robust feature extraction. *IEEE Trans Neural Netw Learn Syst.* 2014 Apr; 25(4):793-805.
- [10] Liu Y, Gao Q, Miao S, Gao X, Nie F, Li Y. A non-greedy algorithm for L1-norm LDA. *IEEE Trans Image Process.* 2017 Feb;26(2):684-95.
- [11] Ye Q, Yang J, Liu F, Zhao C, Ye N, Yin T. L1-norm distance linear discriminant analysis based on an effective iterative algorithm. *IEEE Trans Circuits Syst Video Technol.* 2018 Jan;28(1):114-29.
- [12] Zhang D, Sun Y, Ye Q, Tang J. Recursive discriminative subspace learning with L1-norm distance constraint. *IEEE Trans Cybern.* 2020 May;50(5):2138-51.
- [13] Cevikalp H. Best fitting hyperplanes for classification. *IEEE Trans Pattern Anal Mach Intell.* 2017 Jun;39(6):1076-88.
- [14] Li X, Zhang H, Zhang R, Liu Y, Nie F. Generalized uncorrelated regression with adaptive graph for unsupervised feature selection. *IEEE Trans Neural Netw Learn Syst.* 2019 May;30(5):1587-95.
- [15] Simard P, LeCun Y, Denker JS. Efficient pattern recognition using a new transformation distance. In: Hanson S, Cowan J, Giles C, editors. *Advances in Neural Information Processing Systems*; 1992 Nov 30- Dec 3; San Francisco: Morgan-Kaufmann Press; c1993. p. 50-8.
- [16] Nene SA, Nayar SK, Murase H. 1996 Columbia object image library (COIL-20). Columbia Univ., New York, Tech. Rep. CUCS-005-96.
- [17] Chen X, Yang J, Zhang D, Liang J. Complete large margin linear discriminant analysis using mathematical programming approach. *Pattern Recogn.* 2013 Jun;46(6):1579-94.

# L1-Norm Distance Discriminant Analysis with Multiple Adaptive Graphs and Sample Reconstruction

Guowan SHAO<sup>a,1</sup>, Chunjiang PENG<sup>a</sup>, Wenchu OU<sup>a</sup> and Kai DUAN<sup>b</sup>

<sup>a</sup> College of Mechanical and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411201, Hunan, PR China

<sup>b</sup> Engineering Research Center for Advanced Mine Equipment of Ministry of Education, Hunan University of Science and Technology, Xiangtan 411201, Hunan, PR China

**Abstract.** Linear discriminant analysis (LDA) is sensitive to noise and its performance may decline greatly. Recursive discriminative subspace learning method with an L1-norm distance constraint (RDSL) formulates LDA with the maximum margin criterion and becomes robust to noise by applying L1-norm and slack variables. However, the method only considers inter-class separation and intra-class compactness and ignores the intra-class manifold structure and the global structure of data. In this paper, we present L1-norm distance discriminant analysis with multiple adaptive graphs and sample reconstruction (L1-DDA) to deal with the problem. We use multiple adaptive graphs to preserve intra-class manifold structure and simultaneously apply the sample reconstruction technique to preserve the global structure of data. Moreover, we use an alternating iterative technique to obtain projection vectors. Experimental results on three real databases demonstrate that our method obtains better classification performance than RDSL.

**Keywords.** Discriminant analysis, dimensionality reduction, adaptive graphs, sample reconstruction

## 1. Introduction

High dimensional data are available in pattern recognition and computer vision applications. However, it is difficult to directly use the data for some reasons, such as the curse of dimensionality and high computational complexity [1]. A natural solution to the difficulty is to perform dimensionality reduction or subspace learning for data representation. Many methods are presented to realize the fundamental task under the assumption that the most useful information lies in a latent subspace [2]. Among them, linear discriminant analysis (LDA) [3] and principal component analysis (PCA) [4] are two classical and representative methods for supervised and unsupervised dimensionality reduction, respectively.

---

<sup>1</sup> Corresponding Author, Guowan SHAO, College of Mechanical and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411201, Hunan, PR China; E-mail: gwshao\_ezhou@163.com, 821211624@qq.com.

According to the definition of LDA, it measures within-class and between-class distances in Euclidean space, namely LDA is based on L2-norm. Therefore, LDA is sensitive to noise and its performance may decline greatly. Some methods use L1-norm to measure the distances for enhancing the robustness [5-9]. Rotational invariant LDA (RILDA) [5] uses RI L1-norm or L<sub>2,1</sub>-norm as the metric and constructs a more generalized dimensionality reduction framework. However, the proposed algorithm is not convergent and thus the performance may be further improved. L1-norm LDA [6] extracts local optimal projection vectors by replacing L2-norm with L1-norm based on the formulation of LDA. It avoids the singular problem existing in LDA and obtains better results than RILDA. L1-LDA [7] applies the Bayes optimality theory for excellent performance. However, the iterative algorithm enlarges the Bayes error upper bound and cannot guarantee the optimality. Besides, its robustness cannot be enhanced for the overemphasis on the minimum distance between boundary points and the center of all points. In this situation, non-greedy L1-LDA [8] combines two L1-norm terms in a difference form and gives a non-greedy iterative algorithm, which obtains better performance. L1-norm distance-based effective LDA [9] is similar to non-greedy L1-LDA. However, the two methods are sensitive to initialization for they adopt the Rayleigh quotient as the learning rate.

Recently, a novel recursive discriminative subspace learning method with an L1-norm distance constraint (RDSL) has been proposed to obtain a robust and discriminative subspace [10]. The method formulates LDA with the maximum margin criterion and avoids the sensitiveness to initialization. In addition to L1-norm, RDSL introduces slack variables and thus it is more robust to noise. However, the method only considers inter-class separation and intra-class compactness and ignores the intra-class manifold structure and the global structure of data. It is beneficial to utilize structure information for better performance.

In this paper, we present L1-norm distance discriminant analysis with multiple adaptive graphs and sample reconstruction (L1-DDA) to solve the problem existing in RDSL. We apply Shannon entropy to adaptively construct a graph for each class as a regularization term and thus intra-class structure information may be incorporated into projection vectors. L1-norm is used in the construction of adaptive graphs in order to improve the robustness. Meanwhile, we add a sample reconstruction term to approximately preserve the global structure of data, and sample reconstruction is in the form of cutting squared L2-norm. Finally, we use an alternating iterative technique to optimize the objective function. Experimental results on three publicly available real databases demonstrate that our method obtains better classification performance than RDSL.

## 2. Related Work

In this section, we briefly review RDSL for our method is based on it and then simply explain Shannon entropy which is applied to multiple adaptive graphs. Given a data set  $X = [x_1, x_2, \dots, x_n]$ ,  $x_j \in R^m$ ,  $j = 1, 2, \dots, n$ , and the  $n$  points in the  $m$  dimensional space come from  $c$  classes. Let  $x_{ij}$  denote the  $j$  th point from class  $i$  and  $n_i$  the number of points from class  $i$ . In addition, let  $\bar{x}_i$  denote the mean of class  $i$  and  $\bar{x}$  the mean of all the classes.

## 2.1. Recursive Discriminative Subspace Learning

LDA uses L2-norm to construct an objective function and thus its performance may deteriorate when noisy data are available. In this case, some methods [8-9] adopt L1-norm to cope with the problem. However, the methods are sensitive to initialization. RDSL applies the maximum margin criterion to formulate LDA and avoids the sensitiveness.

RDSL is formulated as [10]

$$\begin{aligned} \min_{w_q, \xi_i} & \sum_{i=1}^c \sum_{j=1}^{n_i} \|w_q^T (x_{ij} - \bar{x}_i)\|_1 + \lambda \sum_{i=1}^c \xi_i \\ \text{s.t. } & \|w_q^T (\bar{x}_i - \bar{x})\|_1 + \xi_i \geq 1 \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, c \end{aligned} \quad (1)$$

where  $w_q$  is the  $q$  th projection vector,  $1 \leq q \leq d$ ,  $d$  is the number of projection vectors, and accordingly the projection matrix  $W = [w_1, w_2, \dots, w_d]$ .  $\xi_i$  is a slack variable, which is introduced to avoid the overfitting and  $\lambda$  is a regularization parameter which can control the generalization or the robustness of RDSL.

Eq. (1) can be transformed into a linear programming problem or a convex quadratic programming problem and then the solution can be obtained by an iterative algorithm [10].

## 2.2. Shannon Entropy

Information theoretic learning techniques have been extensively applied in the pattern recognition field and many methods for dimensionality reduction have involved entropies. Among different types of entropies, Shannon entropy is most widely used since, in general, there is a close-form solution with the entropy.

Shannon entropy is defined as follows [11]:

$$\begin{aligned} H(P) &= -\sum_{i=1}^N p_i \log p_i \\ \text{s.t. } & \sum_{i=1}^N p_i = 1, \quad p_i \geq 0 \end{aligned}$$

where  $P = (p_1, p_2, \dots, p_N)$  is a probability distribution. The definition measures the uncertainty of the distribution. More specifically, the value of  $H(P)$  will increase as the distribution becomes more uniform.

According to the graph theory, the similarity values between samples from the same class tend to the same value. In other words, the values are roughly uniformly distributed. Therefore, the similarity matrix may be optimized by maximizing the Shannon entropy of the distribution. In section 3, we construct an adaptive graph for each class based on the analysis.

### 3. L1-Norm Distance Discriminant Analysis

As formulated in Eq. (1), RDSL considers inter-class separation and intra-class compactness. However, RDSL ignores intra-class structure. It is beneficial to exploit within-class structure information [12] and hence there is room to enhance the performance.

#### 3.1. L1-Norm Distance Discriminant Analysis with Multiple Adaptive Graphs

Since the local geometry structure of data can be modeled by a nearest neighbor graph or a similarity matrix, we use multiple graphs to construct within-class structure. It is difficult to select the number of nearest neighbors of a graph. In order to avoid the difficulty, we intend to construct an adaptive graph for each class. According to the graph theory, the similarity between two samples is large when the samples belong to the same class with high probability. Therefore, it is reasonable to infer that any two samples from the same class have a roughly equal similarity value in most cases. Based on the inference, we use Shannon entropy to adaptively construct similarity matrices.

Our model is formulated as follows:

$$\begin{aligned} \min_{w_q, \xi_i, s_{kj}^i} & \sum_{i=1}^c \sum_{j=1}^{n_i} \|w_q^T (x_{ij} - \bar{x}_i)\|_1 + \sum_{i=1}^c \alpha_i \left( \sum_{k=1}^{n_i} \sum_{j=1}^{n_i} s_{kj}^i \|w_q^T (x_{ik} - x_{ij})\|_1 + \eta_i s_{kj}^i \log s_{kj}^i \right) + \lambda \sum_{i=1}^c \xi_i \\ \text{s.t. } & \|w_q^T (\bar{x}_i - \bar{x})\|_1 + \xi_i \geq 1 \\ & \xi_i \geq 0, \quad \sum_{j=1}^{n_i} s_{kj}^i = 1, \quad s_{kj}^i \geq 0, \quad i = 1, 2, \dots, c, \quad k, j = 1, 2, \dots, n_i \end{aligned} \quad (2)$$

where  $\alpha_i$  and  $\eta_i$  are regularization parameters,  $\alpha_i \geq 0$ ,  $\eta_i \geq 0$ , and  $s_{kj}^i$  is the similarity between the  $k$  th and  $j$  th samples from class  $i$ .

We use L1-norm to measure distances between samples from the same class, which may enhance the robustness of the graphs. We then use multiple adaptive graphs as a regularization term to improve the performance of RDSL. However, intra-class structure is only local geometry structure and our model does not involve global structure. Therefore, the performance may be further improved.

#### 3.2. Improvement Strategy of the Model

The data reconstruction trick is usually applied to preserve the global structure of data [13]. However, the trick may be less than effective when the data are corrupted by noise, especially strong noise. In other words, it is necessary to eliminate the adverse effect of noise on data reconstruction. Inspired by the cutting L1-norm loss function [14], we use cutting squared L2-norm to construct a sample reconstruction term in order to realize the purpose.

The improved model is formulated as follows:

$$\begin{aligned}
& \min_{w_q, u_q, \xi_i, s_{kj}^i} \sum_{i=1}^c \sum_{j=1}^{n_i} \|w_q^T(x_{ij} - \bar{x}_i)\|_1 + \sum_{i=1}^c \alpha_i \left( \sum_{k=1}^{n_i} \sum_{j=1}^{n_i} s_{kj}^i \|w_q^T(x_{ik} - x_{ij})\|_1 + \eta_i s_{kj}^i \log s_{kj}^i \right) \\
& + \beta \sum_{i=1}^c \sum_{j=1}^{n_i} \min \left( h, \|x_{ij} - u_q w_q^T x_{ij}\|_2^2 \right) + \lambda \sum_{i=1}^c \xi_i \\
\text{s.t. } & \|w_q^T(\bar{x}_i - \bar{x})\|_1 + \xi_i \geq 1 \\
& \xi_i \geq 0, \quad \sum_{j=1}^{n_i} s_{kj}^i = 1, \quad s_{kj}^i \geq 0, \quad i = 1, 2, \dots, c, \quad k, j = 1, 2, \dots, n_i
\end{aligned} \tag{3}$$

where  $u_q$  is the  $q$  th reconstruction vector,  $1 \leq q \leq d$ ,  $\beta$  is a regularization parameter,  $\beta \geq 0$ , and  $h$  is a given value,  $h > 0$ .

We do not use cutting L1-norm to measure reconstruction errors for it is easier to optimize the objective function. Although L2-norm is sensitive to noise, it is simple and convenient for optimization. Moreover, cutting squared L2-norm can constrain noisy samples and thus may become robust to noise.

#### 4. Optimization Procedure

Eq. (3) involves the variables:  $w_q$ ,  $u_q$ ,  $s_{kj}^i$  and  $\xi_i$ , and we adopt an alternating iterative method [15] to obtain the solutions.

##### 4.1. Fix $u_q$ and $s_{kj}^i$ and Update $w_q$ and $\xi_i$

When  $u_q$  and  $s_{kj}^i$  are fixed, Eq. (3) reduces to the following form:

$$\begin{aligned}
& \min_{w_q, \xi_i} \sum_{i=1}^c \sum_{j=1}^{n_i} \|w_q^T(x_{ij} - \bar{x}_i)\|_1 + \sum_{i=1}^c \alpha_i \sum_{k=1}^{n_i} \sum_{j=1}^{n_i} s_{kj}^i \|w_q^T(x_{ik} - x_{ij})\|_1 \\
& + \beta \sum_{i=1}^c \sum_{j=1}^{n_i} \min \left( h, \|x_{ij} - u_q w_q^T x_{ij}\|_2^2 \right) + \lambda \sum_{i=1}^c \xi_i \\
\text{s.t. } & \|w_q^T(\bar{x}_i - \bar{x})\|_1 + \xi_i \geq 1 \\
& \xi_i \geq 0, \quad i = 1, 2, \dots, c
\end{aligned} \tag{4}$$

In order to optimize Eq. (4), we first reformulate the first item of the objective function. Clearly,

$$\|w_q^T(x_{ij} - \bar{x}_i)\|_1 = \frac{w_q^T(x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^T w_q}{\|w_q^T(x_{ij} - \bar{x}_i)\|_1}$$

Let  $\rho_{ij} = 1/\|w_q^T(x_{ij} - \bar{x}_i)\|_1$ , then  $\sum_{i=1}^c \sum_{j=1}^{n_i} \|w_q^T(x_{ij} - \bar{x}_i)\|_1 = w_q^T B_{q1} w_q$ , where

$$B_{q1} = \sum_{i=1}^c \sum_{j=1}^{n_i} \rho_{ij} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^T.$$

Similarly, we reformulate the second item of the objective function and

$$\sum_{i=1}^c \alpha_i \sum_{k=1}^{n_i} \sum_{j=1}^{n_i} s_{kj}^i \|w_q^T (x_{ik} - x_{ij})\|_1 = w_q^T B_{q2} w_q$$

$$\text{where } B_{q2} = \sum_{i=1}^c \alpha_i \sum_{k=1}^{n_i} \sum_{j=1}^{n_i} s_{kj}^i \mu_{kj}^i (x_{ik} - x_{ij})(x_{ik} - x_{ij})^T, \mu_{kj}^i = 1/\|w_q^T (x_{ik} - x_{ij})\|_1.$$

Moreover, we reformulate the third item of the objective function, which can be written as

$$\min \left( h, \|x_{ij} - u_q w_q^T x_{ij}\|_2^2 \right) = (1 - \theta_{ij})h + \theta_{ij} \|x_{ij} - u_q w_q^T x_{ij}\|_2^2$$

where  $\theta_{ij} = 1$  if  $\|x_{ij} - u_q w_q^T x_{ij}\|_2^2 < h$ , otherwise  $\theta_{ij} = 0$ . Clearly,

$$\|x_{ij} - u_q w_q^T x_{ij}\|_2^2 = x_{ij}^T x_{ij} - 2x_{ij}^T u_q w_q + w_q^T x_{ij} u_q^T u_q x_{ij} w_q$$

Therefore, we have

$$\sum_{i=1}^c \sum_{j=1}^{n_i} \min \left( h, \|x_{ij} - u_q w_q^T x_{ij}\|_2^2 \right) = b_{q1} + a_{q1}^T w_q + w_q^T G_{q1} w_q$$

where

$$\begin{aligned} b_{q1} &= \sum_{i=1}^c \sum_{j=1}^{n_i} \left( (1 - \theta_{ij})h + \theta_{ij} x_{ij}^T x_{ij} \right) \\ a_{q1} &= -2 \sum_{i=1}^c \sum_{j=1}^{n_i} \theta_{ij} (x_{ij}^T u_q) x_{ij} \\ G_{q1} &= \sum_{i=1}^c \sum_{j=1}^{n_i} \theta_{ij} x_{ij} u_q^T u_q x_{ij}^T \end{aligned}$$

Finally, we reformulate the constrain of the objective function. Clearly,

$$\|w_q^T (\bar{x}_i - \bar{x})\|_1 = \text{sign}(w_q^T (\bar{x}_i - \bar{x})) w_q^T (\bar{x}_i - \bar{x}) = \text{sign}(w_q^T (\bar{x}_i - \bar{x})) (\bar{x}_i - \bar{x})^T w_q$$

where  $\text{sign}(\cdot)$  is the sign function. Let  $e_c$  denote the  $c$  dimensional column vector which includes  $c$  ones, and  $\gamma = [\xi_1, \xi_2, \dots, \xi_c]^T$ . Let  $A_q = [v_1, v_2, \dots, v_c]^T$ , where  $v_i = \text{sign}(w_q^T (\bar{x}_i - \bar{x})) (\bar{x}_i - \bar{x})$ , Then  $\|w_q^T (\bar{x}_i - \bar{x})\|_1 + \xi_i \geq 1$ ,  $i = 1, 2, \dots, c$ , can be reformulated as follows:  $A_q w_q + \gamma \geq e_c$ .

Therefore, Eq. (4) can be reformulated as

$$\begin{aligned} \min_{w_q, \gamma} & w_q^T (B_{q1} + B_{q2} + \beta G_{q1}) w_q + \beta a_{q1}^T w_q + \lambda e_c^T \gamma \\ \text{s.t. } & A_q w_q + \gamma \geq e_c, \gamma \geq 0 \end{aligned} \tag{5}$$

Eq. (5) is a quadratic programming problem, which can be easily solved.

#### 4.2. Fix $w_q$ , $\xi_i$ and $s_{kj}^i$ and Update $u_q$

When  $w_q$ ,  $\xi_i$  and  $s_{kj}^i$  are fixed, Eq. (3) reduces to the following form:

$$\min_{u_q} \sum_{i=1}^c \sum_{j=1}^{n_i} \min \left( h, \|x_{ij} - u_q w_q^T x_{ij}\|_2^2 \right) \quad (6)$$

Eq. (6) can be reformulated as follows:

$$\min_{u_q} g_{q2} u_q^T u_q + a_{q2}^T u_q + b_{q2} \quad (7)$$

where

$$\begin{aligned} b_{q2} &= \sum_{i=1}^c \sum_{j=1}^{n_i} \left( (1 - \theta_{ij}) h + \theta_{ij} x_{ij}^T w_q \right) \\ a_{q2} &= -2 \sum_{i=1}^c \sum_{j=1}^{n_i} \theta_{ij} \left( x_{ij}^T w_q \right) x_{ij} \\ g_{q2} &= \sum_{i=1}^c \sum_{j=1}^{n_i} \theta_{ij} \left( x_{ij}^T w_q \right)^2 \end{aligned}$$

Eq. (7) is also a quadratic programming problem and the solution can be easily obtained.

#### 4.3. Fix $w_q$ , $\xi_i$ and $u_q$ and Update $s_{kj}^i$

When  $w_q$ ,  $\xi_i$  and  $u_q$  are fixed, Eq. (3) reduces to the following form:

$$\begin{aligned} \min_{s_{kj}^i} \sum_{i=1}^c \alpha_i \left( \sum_{k=1}^{n_i} \sum_{j=1}^{n_i} s_{kj}^i \|w_q^T (x_{ik} - x_{ij})\|_1 + \eta_i s_{kj}^i \log s_{kj}^i \right) \\ \text{s.t. } \sum_{j=1}^{n_i} s_{kj}^i = 1, \quad s_{kj}^i \geq 0, \quad i = 1, 2, \dots, c, \quad k, j = 1, 2, \dots, n_i \end{aligned} \quad (8)$$

Eq. (8) can be easily solved and

$$s_{kj}^i = \frac{\exp \left( -\|w_q^T (x_{ik} - x_{ij})\|_1 / \eta_i \right)}{\sum_{j=1}^{n_i} \exp \left( -\|w_q^T (x_{ik} - x_{ij})\|_1 / \eta_i \right)} \quad (9)$$

It is worth noting that  $w_q$  can be initiated by other methods, such as LDA, and obtained by an iterative algorithm. After  $w_q$  is obtained, training samples are updated by the factor  $(I_m - WW^T)$ , where  $I_m$  is the  $m \times m$  dimensional identity matrix [10]. For Eq. (5), since an iterative technique is used, the next value of  $w_q$  can be obtained with the current one. Namely,  $B_{q1}$ ,  $B_{q2}$ ,  $G_{q1}$  and  $a_{q1}$  are also fixed. Similarly, for Eq. (7), the next value of  $u_q$  can be obtained with its current one. The alternating iterative algorithm is summarized in Table 1.

**Table 1.** Algorithm 1. The alternating iterative algorithm

<b>Input:</b> $X$ --- labeled data set, $X \in R^{m \times n}$ ; $\alpha_i, \eta_i, \beta, \lambda$ --- regularization parameters; $d$ --- number of projection vectors ; <b>Output:</b> $W$ --- projection matrix, $W \in R^{m \times d}$ .
<b>Procedure:</b> Initiate the projection matrix $W = \phi$ , where $\phi$ is the empty set; <b>For</b> $q = 1$ <b>to</b> $d$ <b>do</b> Initiate $w_q$ ; Compute $u_q$ according to Eq. (7); Compute $s_{kj}^i$ according to Eq. (9); <b>While</b> not converge <b>do</b> Update $w_q$ according to Eq. (5); Update $u_q$ according to Eq. (7); Update $s_{kj}^i$ according to Eq. (9); <b>End while</b> Let $W = [W, w_q]$ , and normalize and orthogonalize it; Update the data set $X$ , i.e., $X = (I_m - WW^T)X$ ; <b>End for</b>

## 5. Experiments

In this section, we evaluate the classification accuracy of our method, and LDA, coined complete large margin linear discriminant analysis (CLMLDA) [16], RILDA, and RDSL are also run respectively on three real datasets, namely the USPS database [17], the COIL-20 database<sup>2</sup> and the Yale database<sup>3</sup>. For the datasets, the grayscale of images is normalized to  $[0,1]$ . For simplicity, PCA is first used to reduce the dimensionality to  $n - c$  after the images are represented with vectors. For CLMLDA, the optimal values of two parameter are chosen from the set  $[10^{-6}, 10^{-3}, 0, 10^3, 10^6]$ . For RDSL, the optimal value of  $\lambda$  is chosen from the same set. For our method,  $\eta_i$  and  $\beta$  are empirically set to  $10^3$  and 0.1, respectively, and other regularization parameters are chosen from the set. The cutting parameter  $h$  is empirically set and as a result, five percent of training samples with the largest reconstruction errors are eliminated.  $w_q$  is initiated by the solution of LDA and the convergence condition is in accordance with [10]. The nearest neighbor classifier is used to measure classification accuracy and directly used on the preprocessed data by PCA as the baseline method. We repeat all the experiments ten times and obtain average lowest classification error rates.

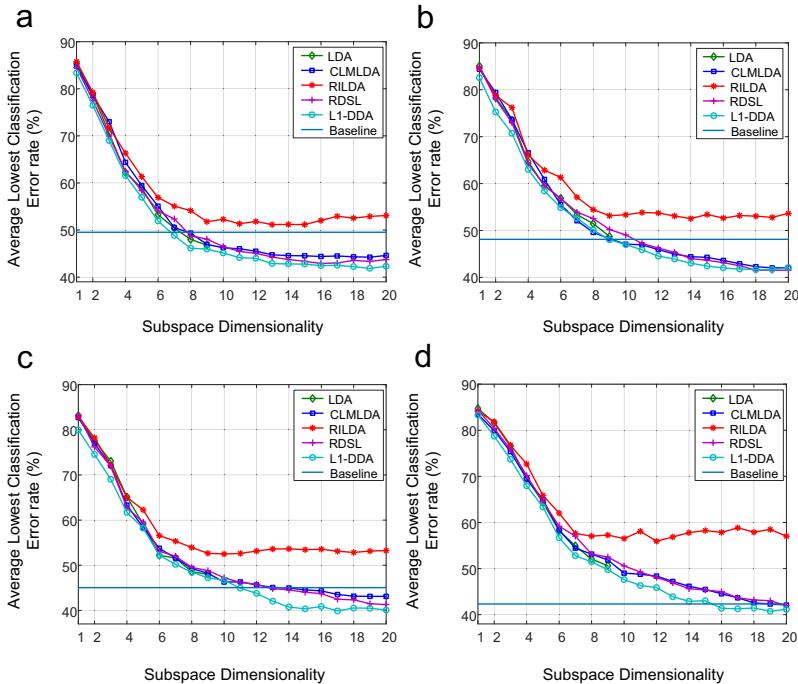
<sup>2</sup> <https://www1.cs.columbia.edu/CAVE/software/softlib/coil-20.php>

<sup>3</sup> <http://vision.ucsd.edu/content/yale-face-database>

### 5.1. Experiments on the USPS Database

The USPS database consists of 9,298 grayscale images of digits from envelopes and their size is  $16 \times 16$  pixels, namely each image can be described as a 256-dimensional vector. For each digit, we randomly select 100 samples and construct a subset for experiments. Furthermore, the images in the subset are artificially contaminated by the Gaussian noise at a random position with the proportion of forty percent. In the experiments, we randomly select 8, 10, 12 and 14 images as training samples, respectively, and the rest images as testing samples.

Average lowest classification error rates with different dimensions are given in Figure 1, which shows that our method slightly outperforms RDSL.

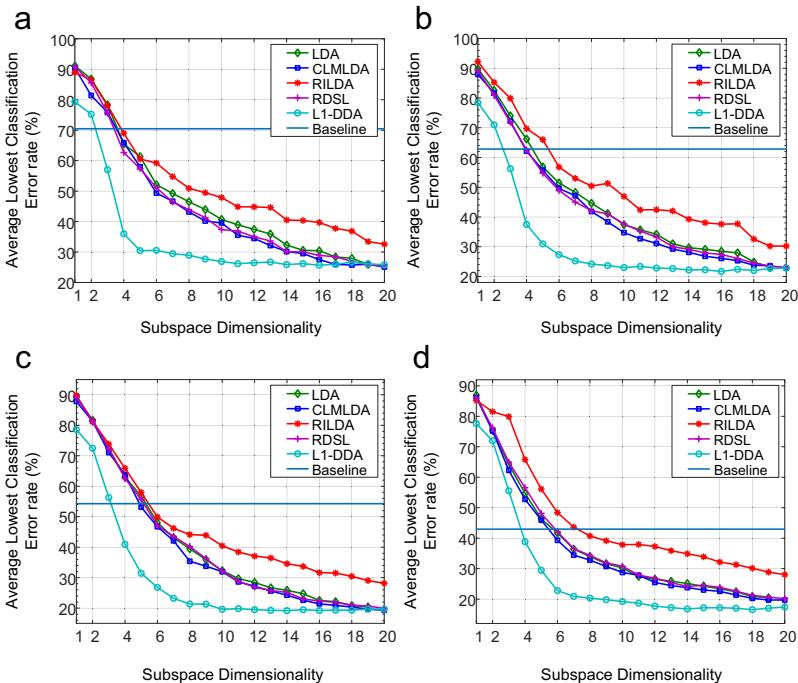


**Figure 1.** Average lowest classification error rates on the USPS database. (a) 8 samples, (b) 10 samples, (c) 12 samples, and (d) 14 samples.

### 5.2. Experiments on the Coil-20 Database

There are 1,440 grayscale images from twenty objects in the COIL-20 database and each object has 72 images which are taken at an interval of five degrees. In the experiments, the images are downsampled and their size becomes  $32 \times 32$  pixels. We then add the Gaussian noise to each image at a random position with the proportion of forty percent. For each object, 6, 8, 10 and 12 images are randomly selected as training samples, respectively, and the rest images are used for testing.

Average lowest classification error rates with different dimensions are shown in Figure 2. From the figure we conclude that the performance of L1-DDA is much better than that of RDSL.



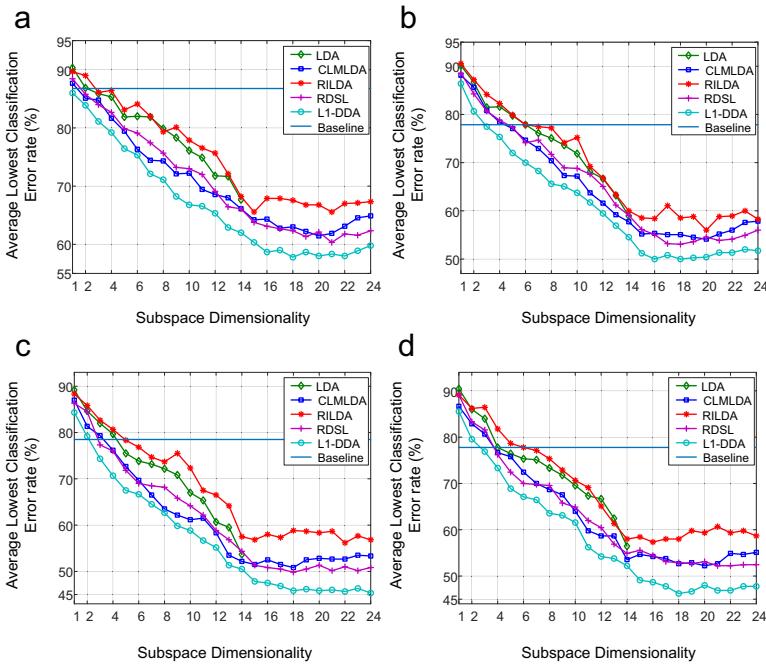
**Figure 2.** Average lowest classification error rates on the COIL-20 database. (a) 6 samples, (b) 8 samples, (c) 10 samples, and (d) 12 samples.

### 5.3. Experiments on the Yale Database

The Yale database consists of 165 grayscale face images and the images are from fifteen subjects and different in three aspects: facial expression, lighting condition and whether or not wearing glasses. In the experiments, we use the processed images whose size is  $32 \times 32$  pixels<sup>4</sup>. Similar to the experiments on the USPS and COIL-20 databases, we use the Gaussian noise to contaminate each image at a random position with the proportion of forty percent. Moreover, we randomly select 5, 6, 7 and 8 images from each subject as training samples, respectively, and the rest ones as testing samples.

Average lowest classification error rates with different dimensions are given in Figure 3. From the experimental results, we can conclude that the proposed method is significantly effective for classification.

<sup>4</sup> <http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html>



**Figure 3.** Average lowest classification error rates on the Yale database. (a) 5 samples, (b) 6 samples, (c) 7 samples, and (d) 8 samples.

## 6. Conclusions

In this paper, we propose L1-norm distance discriminant analysis with multiple adaptive graphs and sample reconstruction (L1-DDA) based on RDSDL. Our method uses multiple adaptive graphs to preserve intra-class manifold structure and applies L1-norm and Shannon entropy to robustly and adaptively construct the graphs. Moreover, it uses the sample reconstruction technique to preserve the global structure of data and cutting squared L2-norm is applied to constrain noisy samples for better robustness. Finally, we use an alternating iterative technique to obtain projection vectors. Experimental results on three real datasets show that our method obtains better classification performance than RDSDL.

## Acknowledgement

The work is supported by Natural Science Foundation of Hunan Province (Grant No.2018JJ2133).

## References

- [1] Duda RO, Hart PE, Stork DG, editors. Pattern classification. 2nd ed. NewYork: Wiley Press; c2005.

- [2] Hastie T, Tibshirani R, Friedman J, editors. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. Berlin: Springer Press; c2001.
- [3] Belhumeur PN, Hespanha JP, Kriegman DJ. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell. 1997 Jul;19(7):711-20.
- [4] Jolliffe IT, editors. Principal component analysis. 2nd ed. New York: Springer-Verlag Press; c2002.
- [5] Lai Z, Xu Y, Yang J, Shen L, Zhang D. Rotational invariant dimensionality reduction algorithms. IEEE Trans Cybern. 2017 Nov;47(11):3733-46.
- [6] Zhong F, Zhang J. Linear discriminant analysis based on L1-norm maximization. IEEE Trans Image Process. 2013 Aug;22(8):3018-27.
- [7] Zheng W, Lin Z, Wang H. L1-norm kernel discriminant analysis via Bayes error bound optimization for robust feature extraction. IEEE Trans Neural Netw Learn Syst. 2014 Apr;25(4):793-805.
- [8] Liu Y, Gao Q, Miao S, Gao X, Nie F, Li Y. A non-greedy algorithm for L1-norm LDA. IEEE Trans Image Process. 2017 Feb;26(2):684-95.
- [9] Ye Q, Yang J, Liu F, Zhao C, Ye N, Yin T. L1-norm distance linear discriminant analysis based on an effective iterative algorithm. IEEE Trans Circuits Syst Video Technol. 2018 Jan;28(1):114-29.
- [10] Zhang D, Sun Y, Ye Q, Tang J. Recursive discriminative subspace learning with L1-norm distance constraint. IEEE Trans Cybern. 2020 May;50(5):2138-51.
- [11] Fukunaga K, editors. Introduction to statistical pattern recognition. New York: Academic Press; c1991.
- [12] Hu L, Zhang W. Orthogonal neighborhood preserving discriminant analysis with patch embedding for face recognition. Pattern Recogn. 2020 Oct;106(10):1579-94.
- [13] Wen J, Han N, Fang X, Fei L, Yan K, Zhan S. Low-rank preserving projection via graph regularized reconstruction. IEEE Trans Cybern. 2019 Apr;49(4):1279-91.
- [14] Cevikalp H. Best fitting hyperplanes for classification. IEEE Trans Pattern Anal Mach Intell. 2017 Jun;39(6):1076-88.
- [15] Li X, Zhang H, Zhang R, Liu Y, Nie F. Generalized uncorrelated regression with adaptive graph for unsupervised feature selection. IEEE Trans Neural Netw Learn Syst. 2019 May;30(5):1587-95.
- [16] Chen X, Yang J, Zhang D, Liang J. Complete large margin linear discriminant analysis using mathematical programming approach. Pattern Recogn. 2013 Jun;46(6):1579-94.
- [17] Simard P, LeCun Y, Denker JS. Efficient pattern recognition using a new transformation distance. In: Hanson S, Cowan J, Giles C, editors. Advances in Neural Information Processing Systems; 1992 Nov 30- Dec 3; San Francisco: Morgan-Kaufmann Press; c1993. p. 50-8.

# Dist Frequent Next Neighbours: A Distributed Galois Lattice Algorithm for Frequent Closed Itemsets Extraction

Naomie Sandra NOUMI SANDJI<sup>a,1</sup>, Djamal Abdoul Nasser SECK<sup>a</sup>

<sup>a</sup>*Department of Mathematics and Computer Science, Faculty of Science and Technology, Cheikh Anta Diop University of Dakar, Senegal*

**Abstract.** The general purpose of this paper is to propose a distributed version of frequent closed itemsets extraction in the context of big data. The goal is to have good performances of frequent closed itemsets extraction as frequent closed itemsets are bases for frequent itemsets. To achieve this goal, we have extended the Galois lattice technique (or concept lattice) in this context. Indeed, Galois lattices are an efficient alternative for extracting closed itemsets which are interesting approaches for generating frequent itemsets. Thus we proposed Dist Frequent Next Neighbour which is a distributed version of the Frequent Next Neighbour concept lattice construction algorithm, which considerably reduces the extraction time by parallelizing the computation of frequent concepts (closed itemsets).

**Keywords.** frequent itemset, association rules, closed itemset, Galois lattice, big data, data mining.

## 1. Introduction

Association rule mining is a data mining task introduced by Agrawal et al. in 1993 [1] that discovers meaningful relationships between attributes according to their associations in databases. It consists of two steps: the discovery of frequent itemsets and the generation of association rules. The discovery of frequent itemsets, which is the most important step, consists in discovering sets of attributes that appear very often with the same groups of values on a large part of the data. There are several areas of application. In the **commercial sector** [1], [2], the analysis of the transaction databases allows to determine the items that are often bought together and thus to highlight the causal relationships between them, which then allows to elaborate a good commercial planning as to the location of the items in the shelves. In the **telecommunication sector**, they are used for example for the prediction of incidents in remote maintenance processes, in order to limit the costs of manual interventions and to improve the quality of service[3]. In the **medical field** [4] ,organizations (laboratories, hospitals, etc.) can search frequent

---

<sup>1</sup>Corresponding Author: Naomie Sandra NOUMI SANDJI, Cheikh Anta DIop University of Dakar, Senegal, B.P.:5005 Dakar-Senegal; E-mail:noumisandji03@gmail.com.

sets in patient databases to determine symptoms that precede an illness, or a treatment to be provided. Several algorithms for frequent itemsets extraction have been proposed, including the Apriori algorithm [5], which is one of the first algorithms proposed, and its improved variants such as ECLAT [6], Fp-growth [7], etc. These algorithms sequentially analyze the databases to extract frequent itemsets. With the explosion of data favored by the development of the internet and data collection and storage tools, the size of transaction databases becomes very important and the extraction task becomes a complex and computationally expensive task, as it requires several scans of the transaction database for the calculation of the support of the candidate itemsets in order to generate the frequent ones. Next to these algorithms there are others based on frequent closed itemsets such as CLOSE [8], CHARM [9], CLOSET [10], etc. Frequent closed itemsets constitute reduced bases allowing to find all frequent itemsets, therefore reducing the set of candidates to be processed and the number of scans in the databases. Thus the extraction of frequent closed itemsets is more advantageous in the context of large volumes of data, as they allow the improvement of the computation time. In this paper we propose **Dist Frequent Next Neighbour**, which is a distributed version of the Frequent Next Neighbour algorithm for extracting frequent closed itemsets from Galois (or concept) lattices. We implemented this algorithm with the Apache Spark framework. This proposed algorithm allows the efficient extraction of frequent closed itemsets by optimizing the closure computation and the support computation of closed itemsets. The plan of the paper is organized as follows: after presenting some definitions in section 2, a literature review is presented in section 3. Then section 4 presents in detail the Dist frequent Next Neighbour algorithm; section 5 evaluates the performances of the Dist Frequent Next Neighbour algorithm and compares them to those of the Spark-Apriori algorithm [11] and finally section 6 to conclude the paper.

## 2. Some Definitions of the Term

In this section we present some basic notions for the good understanding of the document.

A **Formal context** also called transaction base, is a triplet  $\mathbf{K} = (\mathbf{O}, \mathbf{A}, \mathbf{I})$ , where  $\mathbf{O}$  is the set of objects or transactions,  $\mathbf{A}$  is the set of attributes and  $\mathbf{I}$  is the incidence relation between  $\mathbf{O}$  and  $\mathbf{A}$  ( $\mathbf{O}$  and  $\mathbf{A}$  are two disjoint sets). We write  $\mathbf{oIa}_i$  or  $(\mathbf{o}, a_i) \in \mathbf{I}$  to mean that the object  $\mathbf{o}$  has for attribute  $a_i$ . It is represented in the form of an array in which the objects are in row and the attributes in column, we use 1 or X to represent the incidence relation  $\mathbf{oIa}_i$ . If an object  $\mathbf{o}$  does not have an attribute  $a_i$  we leave the box empty or we put 0 in the box. Let  $\mathbf{O} = \{1, 2, 3, 4, 5, 6, 7\}$  the set of objects and  $\mathbf{A} = \{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{e}, \mathbf{f}\}$  the set of attributes, the table 1 below presents an example of formal context.

**Table 1.** Example of a formal context[12]

Objects	Attributes					
	a	b	c	d	e	f
1	X	X	X	X	X	
2	X		X			
3		X	X			
4	X	X		X		
5			X	X		
6		X		X		
7				X		X

A **item** or **item** or **attribute**, is any element belonging to a finite set of distinct elements.  $I = \{x_1, x_2, \dots, x_m\}$ .

**Example:** In the table 1, the finite set of elements  $I = \{a, b, c, d, e, f\}$  contains 5 items a, b, c, d, e, f.

An **itemset** is a set of items or any subset of items of  $I$ . The number of items of an itemset constitutes its length, an itemset containing  $k$  items is called a **k-itemsets**

The **support** of an itemset is the percentage of transaction instances that contain the itemset. It allows to measure the interest of the itemset, mathematically the support of an itemset  $X$  is defined by :

**Support(X) = Card** ( $\{t_i/X \subseteq t_i, t_i \in T\}$ ). Where  $\text{Card}(A)$  is the cardinal of the set A.

A **frequent itemset** is an itemset whose support verifies the fixed support threshold  $y$ . In other words the itemset X is frequent if and only if  $\text{support}(X) \geq y$ .

A **Galois correspondence** [13] is a **derivation operation** which allows to establish the link (the correspondence) between objects and attributes of a context **K**. It can be defined from two functions **f** and **g** :

- Soit  $X \subseteq O, f : P(O) \rightarrow P(A), f(X) = \{ a \in A / \forall o \in X, o \text{la} \}$
- Soit  $Y \subseteq A, g : P(A) \rightarrow P(O), g(Y) = \{ o \in O / \forall a \in Y, o \text{la} \}$

**Example** in the table 1,  $f(\{3,5\}) = \{c\}$  and  $g(\{a,c\}) = \{1,2\}$  which means that the set of attributes  $\{a,c\}$  has as a set of common objects  $\{1,2\}$ . In the same way, the set of objects  $\{3,5\}$  has in common the attribute  $\{c\}$ .

The **Closure of sets** [13] of a set S is the smallest superset containing this set. The two functions f and g will be used to compute the closure of X ( $X \subseteq O$ ) and the closure of Y ( $Y \subseteq A$ ) representing respectively a subset of objects and a subset of attributes. To do this we compose the functions f and g as follows:

- $X'' = g(f(X))$  : closing operator on objects
- $Y'' = f(g(Y))$  : closing operator on attributes.

A set is said to be closed if it is equal to its closure. Thus X is closed if  $X = X''$  (respectively Y is closed if  $Y = Y''$ ).

A closed itemset is said to be **frequent** if its support verifies the minimum fixed support

A **Formal concept** is a couple  $(O_1, A_1)$ , with  $O_1 \subseteq O, A_1 \subseteq A, O_1 = A'_1$ , and  $A_1 = O'_1$ .  $O_1$  is the extension (object) of  $A_1$  and  $A_1$  is the intention (attribute) of  $O_1$ .

$O'_1, A'_1$  correspond respectively to the derivation operations f ( $O_1$ ) and g ( $A_1$ )

**Example :**  $(\{1,2,3,5\}, \{c\})$  and  $(\{1,2\}, \{a,c\})$  are both formal concepts.

A **Concept lattice** is a pair  $(S, \leq)$ , where:  $\leq$  is an **order relation** on the set S. That is to say a binary relation which verify the following properties:

1. reflexivity : for all  $x \in S$  we have  $xR_x$
2. antisymmetry: for all  $x, y \in S$  we have ( $xRy$  and  $yRx$ ) involves  $x = y$
3. transitivity: for all  $x, y, z \in S$  we have ( $xRy$  and  $yRz$ ) involves  $xRz$

### 3. State of Art

Several algorithms have been proposed to solve the problem of itemset extraction in large volumes of data [14,15,16,17,18,19]. The majority of these algorithms use techniques, tools and strategies to solve the problems of generating a large number of candidates in the candidate itemset generation phase, efficient support counting in order to improve retrieval performance. They are also much more based on frequent itemsets extraction algorithms such as Apriori [5], ECLAT [6] and Fp-Growth [7]. In this paper we propose the Dist Frequent Next Neighbours algorithm for frequent closed itemsets extraction which improves and speeds up the frequent itemsets extraction process as frequent closed itemsets form a basis for frequent itemsets.

### 4. The Dist Frequent Next Neighbours Algorithm

In this section we will first present the Frequent Next Neighbours algorithm, then the design of the Dist Frequent Next Neighbours distributed algorithm followed by its architecture and implementation.

#### 4.1. The Frequent Next Neighbours Algorithm

The **Frequent Next Neighbours** [20] algorithm used to address the problem of determining only frequent concepts, rather than all concepts. We would like to recall that a **Frequent Concept** is a tuple (**extension, intention**) with the intention being a **Frequent Closed Topic**. The Next Frequent Neighbour algorithm is a top-down breadth-first search algorithm (we start with the parent concept and generate its descendants (direct successors). It is formalized by the function **Find Frequent Lower Neighbours** which computes and/or lists all the frequent lower neighbors (successors/children) of each concept starting with the parent concept. The Frequent Next Neighbours sequential algorithm operates in five steps that we have identified ourselves, namely:

**step 1:** The combinations of the intentions. In this step, combinations (union) are made between the attribute (intention) of the current concept and each attribute of the set of attributes of the context, excluding the attribute of the current concept. The objective is to form potential intentions (or candidate intentions) **Y1**, for the calculation of the lower neighboring concepts of the current concept.

**Step 2:** The support check. In this step, the support of each of the combinations obtained in step 1 is computed and we retain those which have a support higher than or equal to the fixed minimum support.

**Step 3:** The calculation of the neighboring concept. For each combination **Y1** retained in step 2 we compute the couple (**Y1'**, **Y1''**), where **Y1'** and **Y1''** are the first and second derivatives which correspond respectively to the Galois connection functions **g** and **fog**, **g** which associates the attributes with their objects and **fog** which is the closure operator on the attributes.

**Step 4** : The maximality check. This step makes it possible to check the relation of coverage (direct link) between the parent concept and the calculated neighbour concept (child concept) .

**Step 5:** The existence check. This last step allows to be reassured of the uniqueness of a generated frequent concept. Generally, the first three steps are grouped into a single step called the generation of frequent neighbour concepts of a concept.

#### 4.2. Design of the Dist Frequent Next Neighbours Algorithm

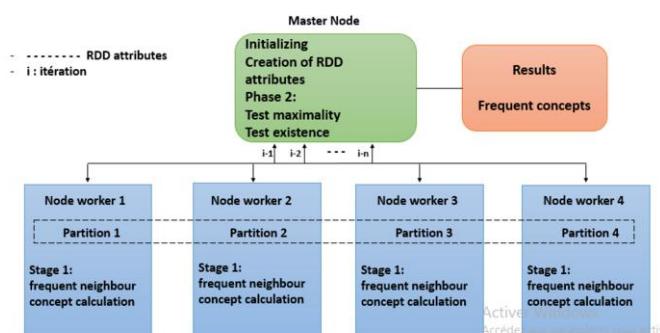
In order to be able to design a parallel/distributed architecture of the Frequent Next Neighbour concept generation algorithm, we first identified the independent actions of the algorithm that can participate to the reduction of the execution time. We thus identified the step of generating the frequent neighbour concepts of a concept as an independent step, because it can be carried out in a completely independent way. This step which is the most complex step can be executed by several nodes simultaneously. And the maximality and existence checks are delegated to a main node, because they require the generation of all concepts by each node before being performed. Thus our distributed approach is carried out in two phases:

**Stage 1 : Parallel computation of frequent neighbour concepts.** It is carried out in three stages including the combinations of the intentions, the support control and the calculation of the frequent neighboring concepts of a concept.

**Stage 2 : The generation of frequent formal concepts (frequent closed itemsets)** This is done in two steps, the maximality check and the existence check.

#### 4.3. Architecture of the Dist Frequent Next Neighbours Algorithm

To improve the performance of the frequent closed itemset extraction process, the Dist Frequent Next Neighbours algorithm is based on the logical distribution of the search space and a parallel execution of the generation of frequent neighbour concepts of a frequent concept which are frequent closed itemset. Indeed, we make a distribution of the set of attributes and on each distribution we have distinct combinations of attributes that constitute subsets of the search space. Thus the architecture of our distributed approach is given by the figure 1 below:



**Figure 1.** Architecture of the Frequent Next Neighbour distributed algorithm.

#### 4.4. Implementation on Spark

This subsection presents in detail the different steps of the implementation of the distributed algorithm Dist Frequent Next Neighbours in the Spark environment via the **pyspark** python API.

##### **Stage I: Parallel calculation of frequent neighbour concepts**

The algorithm starts by creating the object **SparkContext**, whose role is to coordinate Spark applications, running as independent sets of processes on a cluster. It takes as input the formal context file, the structure of this file is such that objects and attributes are separated by the **space** character . The input file is initially loaded into a Spark RDD. The use of RDDs makes it possible to make the most of the resources available at the cluster level and also take advantage of fault tolerance. To facilitate access to these data sets, broadcast variables (command **broadcast()**) were used to store the data dictionaries in the cache of each node-slave. Next we start by creating our parent concept; we create an RDD that contains the set of all context attributes, excluding the attributes of the current concept. Initially our current concept is the parent concept. This rdd will allow us to perform the parallel calculation of the frequent neighbourconcepts of a concept.

##### **Stage II: The generation of frequent formal concepts (frequent closed itemset)**

After each iteration of generating the frequent neighbour concepts of a concept, the results are sent to the master node by the spark action **collect()** for maximality and existence testing to generate the relevant and unique frequent concepts (frequent closed itemset). stages I and II are repeated for all these generated frequent concepts until there are no more frequent concepts.

## 5. Experimental Assessment

In this section, we analyze the performance of our algorithm by analyzing the extraction time of frequent (closed) itemset compared to those of the distributed Apriori algorithm on Spark (Spark-Apriori [11]). For this purpose, a series of several tests is performed by varying the value of the minimal support.

### 5.1. Data Sets

We have used for our experiments the following databases : Mushrooms, Retails, T40I10D100K. These databases are available on the internet. They are distinguished by their type. We have the **dense databases** which better reflect the real transaction databases, they produce a significant number of frequent itemset of rather large size, and this even for high values of the support. They are also characterized by a large number of attributes per object (transaction), and a limited number of distinct attributes. We also have the **sparse databases** which most often reflect the synthetic transaction bases, characterized by a few items per object and a large number of distinct attributes. The Mushrooms databases contain information about mushrooms, they contain 8124 transactions with an average size of 23 attributes per object and 115 items corresponding to the characteristics of mushrooms in total. Retails and T40I10D100K are synthetic databases built according to the properties of sales data. They contain respectively 88162 and 100 000 transactions with an average size of 15 attributes per object for Retails and 40 attributes

for T40I10D100K. The table 2 lists the characteristics of these databases we used for our tests.

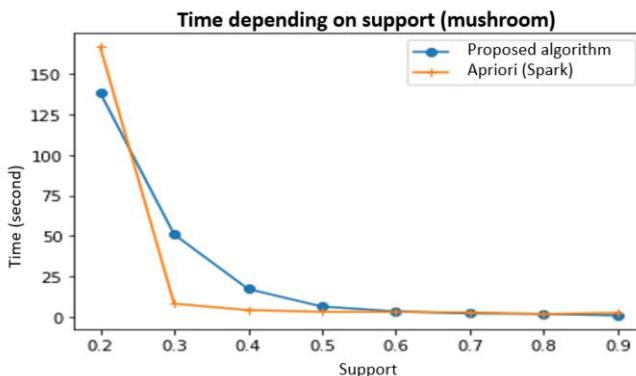
**Table 2.** Characteristics of transactional test databases

Basics	Type	Number of items	Number of transactions
Mushrooms	Dense	115	8124
T40I10D100K	Sparse	1000	100 000
Retails	Sparse	16469	88 162

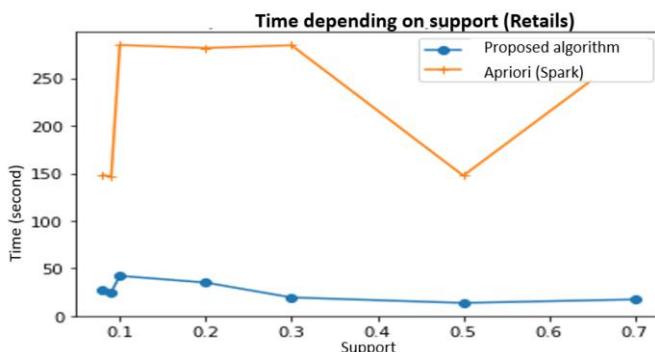
For the performance study we used a cluster on amazon's cloud with 3 nodes whose characteristics are: 8 virtual cores, 32GB of RAM and 128GB SSD.

### 5.2. Performance Evaluation

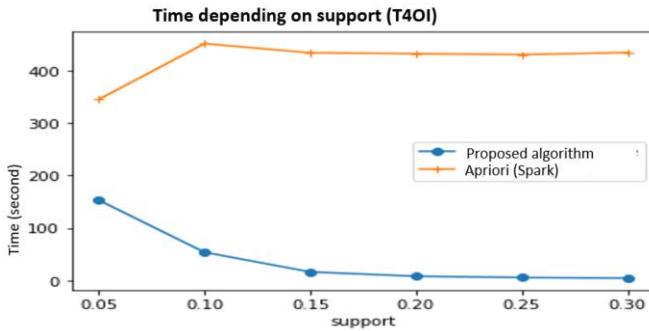
The results obtained are represented by the figures 2, 3 and 4.



**Figure 2.** Variation of the extraction time of the Spark-Apriori algorithm and the proposed algorithm for the Mushrooms database as a function of the support thresholds.



**Figure 3.** Variation of the extraction time of the Spark-Apriori algorithm and the proposed algorithm for the database Retails as a function of support thresholds.



**Figure 4.** Variation of the extraction time of the Spark-Apriori algorithm and the proposed algorithm for the T40I10D100K base as a function of the support thresholds.

### Interpretations:

Reading of the figures 2-4. We could know that :

- For the Mushrooms dataset, the response times of the Apriori distributed algorithm are much lower than those of the Dist Frequent Next Neighbours algorithm for support thresholds of 0.3, 0.4, and 0.5. This can be explained by the fact that the Mushrooms dataset is dense and correlated. Then, we have a large number of frequent itemsets generated during iterations, which is less advantageous for our proposed algorithm, as it performs more operations than the distributed Apriori algorithm. And for support thresholds of 0.7, 0.8 and 0.9, the response times of the two algorithms tend to be identical. This is due to the fact that for these minimum support values the number of frequent itemsets generated at each iteration is approximately the same as the number of closed itemsets generated. As a result, the search spaces are almost identical.
- For the Retails and T40I10D100K datasets, the response times of the Dist Frequent Next Neighbours algorithm are much lower than those of the Apriori distributed algorithm whatever the value of the minimum support set. We observe a degradation of the performance of the distributed Apriori algorithm.

## 6. Conclusion and Future Works

We presented a new approach for frequent itemsets extraction based on closed itemsets. The proposed algorithm Dist Frequent Next Neighbour is an extension of the Frequent Next Neighbour algorithm to big data. The fact that frequent closed itemsets are by nature reduced sets of frequent itemsets has allowed to obtain a significant reduction of the search space to be explored (reduction of the number of candidates to be considered at each iteration). The experimental results show an efficiency compared to the Spark-Apriori algorithm. However, we note that the concept computation phase is the most computationally intensive phase of our algorithm; this is explained by the intersection operations performed between the sets of objects and attributes which contain more and more elements. In our future work, we plan to improve the concept computation by optimizing or dispensing with the intersection computation between the sets of objects and attributes which is very consuming in terms of memory space.

## References

- [1] Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases. In : Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data;1993. p. 207-216.
- [2] White EN. The stock market boom and crash of 1929 revisited. Journal of Economic Perspectives.1990;4(2):67-83.
- [3] Ali K, Manganaris S, Srikant R. Partial Classification Using Association Rules. In : KDD; 1997; p. 115-118.
- [4] Pazzani M, Mani S, Shankle RW. Comprehensible knowledge discovery in databases. In : Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society; 1997; p. 596-601.
- [5] Agrawal R, Srikant R. Fast algorithms for mining association rules. In : Proc. 20th Int. Conf. Very Large Data Bases, VLDB; 1994; p. 487-499.
- [6] Zaki MJ, Parthasarathy S, Ogihsara M, Li W. Parallel algorithms for discovery of association rules. Data Mnning and Knowledge Discovery. 1997;1(4):343-373.
- [7] Han J, Pei J, Yin Y, Mao R. Mining frequent patterns without candidate generation: A frequent-pattern tree approach. Data Mining and Knowledge Discovery. 2004;8(1):53-87.
- [8] Pasquier N, Bastide Y, Taouil R, Lakhal L. Efficient mining of association rules using closed itemset lattices. Information Systems. 1999;24(1):25-49.
- [9] Zaki MJ, Hsiao CJ. CHARM: An efficient algorithm for closed itemset mining. In : Proceedings of the 2002 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics; 2002; p. 457-473.
- [10] Pei J, HAN J, MAO R. Closet: An efficient algorithm for mining frequent closed itemsets. In : ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery; 2000; p. 21-30.
- [11] Zhang943. Spark apriori [Online]. 2018. Available from: <https://github.com/zhang943/Spark-Apriori.git>.
- [12] Jatteau G. APPROXIMATION DU TREILLIS DE CONCEPTS POUR LA FOUILLE DE DONNÉES. 2005.
- [13] Ganter B, Wille R. Contextual attribute logic. In : International Conference on Conceptual Structures. Springer, Berlin, Heidelberg; 1999; p. 377-388.
- [14] Li H, Wang Y, Zhang D, Chang EY. Pfp: parallel fp-growth for query recommendation. In : Proceedings of the 2008 ACM Conference on Recommender Systems; 2008; p. 107-114.
- [15] Yahya O, Hegazy O, Ezat E. An efficient implementation of apriori algorithm based on hadoop-mapreduce model. International Journal of Reviews in Computing. 2012;12:59-67.
- [16] Moens S, Aksehirli E, Goethals B. Frequent itemset mining for big data. In : 2013 IEEE International Conference on Big Data. IEEE; 2013; p. 111-118.
- [17] Qiu H, Gu R, Yuan C, Huang Y. Yafim: a parallel frequent itemset mining algorithm with spark. In : 2014 IEEE International Parallel & Distributed Processing Symposium Workshops. IEEE; 2014; p. 1664-1671.
- [18] Zhang F, Liu M, Gui F, Shen W, Shami A, Ma Y. A distributed frequent itemset mining algorithm using Spark for Big Data analytics. Cluster Computing. 2015;18(4):1493-1501.
- [19] Rathee S, Kashyap A. Adaptive-Miner: an efficient distributed association rule mining algorithm on Spark. Journal of Big Data. 2018;5(1):1-17.
- [20] Carpineto C, Romano G. Concept data analysis: Theory and applications: John Wiley & Sons; 2004.

# Monitoring and Diagnosis System of Downhole Tubing Leakage

Yunpeng YANG<sup>a,1</sup>, Jianchun FAN<sup>b</sup>, Di LIU<sup>a,b</sup>, Fanfan MA<sup>b</sup>

<sup>a</sup>*China University of Petroleum-Beijing, China*

<sup>b</sup>*The Third Research Institute of China Electronics Technology Group Corporation,  
China*

**Abstract.** The downhole tubing in a gas well is affected by many factors such as high pressure erosion, gas lift operation, sand production at the bottom of the well and engineering construction, etc., which can easily lead to leakage of the threaded joints of the tubing and the pipe body, and the leaked natural gas will invade Annulus, making the annulus under pressure. The annular pressure caused by oil pipe leakage is a major safety hazard in oil and gas production. Therefore, the accurate diagnosis the degree of leakage of downhole tubing is of great significance to preventing the occurrence of production accidents effectively. To this end, a set of downhole tubing leak monitoring and diagnosis system has been developed by integrating fluid monitoring, acoustic wave detection and tracer detection technology, and the developed tubing leak monitoring and diagnosis system was used for leak detection tests on offshore platforms. The test results show that the developed tubing leakage monitoring and diagnosis system can meet the need of offshore gas well diagnosis, and realize the holographic diagnosis of the leakage degree of the downhole tubing without moving the downhole tubing string.

**Keywords.** Gas well; tubing leakage; monitoring; diagnosis system; holographic diagnosis

## 1. Introduction

The pressure in annulus caused by tubing leakage continues to threaten the safe production of gas well. Active measures should be taken to diagnose the leakage status. At present, major domestic and foreign oil and gas group companies and related scholars mainly adopt two methods: downhole detection and wellhead detection. Downhole detection is a direct and effective detection method. Common downhole detection technologies include: temperature logging, noise logging, ultrasonic logging, electromagnetic logging, etc. These methods have a large number of application cases in logging, and excellent results have been achieved. But downhole testing technology requires the testing equipment to be placed in the downhole tubing or casing, the production well needs to be shut down, or even the downhole tubing string needs to be moved, which will inevitably bring huge economic losses and safety risks. Therefore, the development of tubing Leakage wellhead detection technology is very necessary, and it is also an important direction of scientific research for major oil and gas group companies and related universities. Research on tubing leakage wellhead detection

---

<sup>1</sup> Corresponding Author: Yunpeng Yang, China University of Petroleum, Beijing, China; E-mail yy1914624371@163.com

technology as early as 1995, RJ Hwang et al. studied the use of changes in crude oil composition during different production periods to determine the leakage between different completion strings of oil wells. However, the implementation of this method requires years of experience. Based on crude oil analysis data, rapid detection is not possible [1]. With the in-depth study of the annulus pressure problem, new achievements have been made in the field of downhole leakage detection. In 2000, Adam T. Bourgoyn et al. proposed to sample annulus fluid and analyze its hydrocarbon composition when the annulus was depressurized, and then compared it with the composition of known pay zones to determine whether the annulus pressure was caused by leakage [2]. They also believe that the growth rate of annulus pressure is related to the size and location of downhole leaks. The API RP90 standard introduces a method for judging the general position of tubing leakage based on the shut-in pressure test: close the downhole safety valve (surface controlled subsurface safety valve, SCSSV), release the pressure on the upper part of SCSSV, if the pressure of A annulus drops, then the leakage point is located in the upper part of the SCSSV, and the leakage in the lower part of the SCSSV can be judged by lowering the bridge plug into any position in the tubing [3]. In 2012, Zhu H. et al. studied the mechanism and prediction model of pressure rise in the "A" annulus in CO<sub>2</sub> injection wells, they proposed a method for predicting the location of tubing leaks based on the principle of pressure balance. They believed that when the annulus pressure reached a stable level, The internal and external pressures at the leak point are balanced, and the iterative calculation process of the leak point position is given [4]. Later, I. Tarmoom et al. proposed the measurement of pressure, flow, fluid composition and other parameters in the comprehensive annulus pressure relief process to estimate the annulus leakage rate, and developed a set of annulus pressure relief rate measurement system in Adma-Opc. The field measurement application [5]. The Norwegian ScanWell company has also developed a monitoring system that integrates the leak detection system (LMS), sonic logging system (AMS), tracer detection system (TMS), component analysis and other functions. Among them, LMS is used to measure downhole wells. The leakage rate of gas, liquid or gas-liquid two-phase at the barrier failure position. AMS is used to confirm abnormal conditions of tubing and annulus and to detect liquid level changes. TMS is used to confirm the leakage depth of tubing and casing, which can realize the Diagnose the leakage status of downhole tubing and casing under the condition of column. Fan Jianchun and his colleagues at the China University of Petroleum have also made significant contributions to the study of annulus pressure, especially in the identification of acoustic leakage, and have proposed a method for the location of downhole casing leakage based on acoustic leakage.

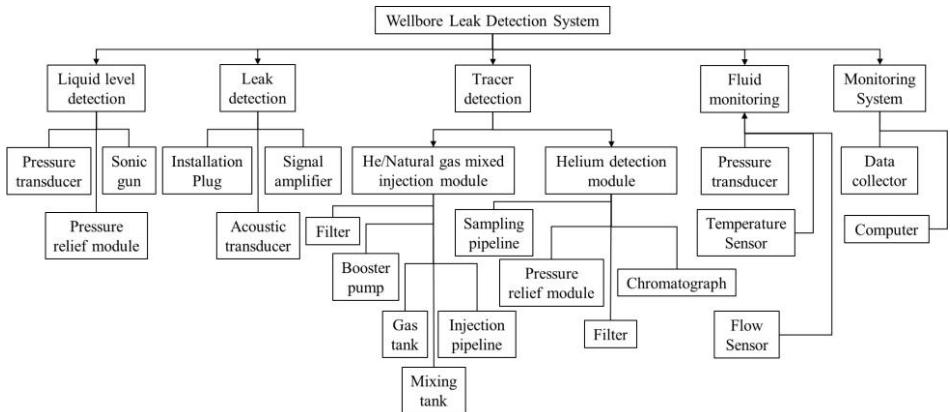
Based on the characteristics of gas wellbore structure and string leakage acoustic waves, a set of downhole tubing leak detection system is designed that integrates multi-fluid parameters, acoustic wave method and tracer detection. Among them, acoustic wave method is used to detect annulus fluid level and pipe leakage. The column leakage point is located, and the tracer method is used to estimate the equivalent size of the leakage hole.

## 2. The composition and principle of the system

### 2.1. The Composition of the system

As shown in Figure 1: The wellbore leakage monitoring system is mainly composed of fluid monitoring and signal acquisition, liquid level detection, leak detection, and tracer detection systems.

- Fluid monitoring: Use pipelines to introduce the annular air into the system, and monitor the gas status during the depressurization process. The main monitoring parameters include the flow rate, pressure and temperature in the pressure relief pipeline, and at the same time sampling and analyzing its composition at the pressure relief port.
- Liquid level detection: Estimating the depth of the annulus liquid level by testing the liquid level echo signal.
- Leak point detection: According to the structure of the Christmas tree tubing four-way, the flow channel of the four-way gate valve is connected to the annulus, and the acoustic wave sensor can be placed in it. When the gate valve is opened, the acoustic wave in the annulus can be received signal.
- Tracer detection: It mainly includes tracer injection and detection. The tracer is injected from the annulus, and the tubing and adjacent annulus outlets are inspected, and the gas well production is not affected during the inspection process. Non-polluting helium gas is selected as the tracer source. Considering that the annulus pressure is generally high, the tracer needs to be pressurized before it can be injected successfully, and the gas produced needs to be depressurized before the test is performed. The whole process includes three modules: helium pressurization-gas injection-helium detection.



**Figure 1.** The composition of the downhole tubing leak monitoring and diagnosis system

### 2.2. Principle

According to APR RP90, the causes of pressure in the oil jacket annulus include thermal effects, gas lift residual pressure, oil pipe leakage and production casing leakage [3]. After confirming the existence of the leakage, it is necessary to determine the specific

leakage source. As shown in Figure 2, a comprehensive diagnosis and analysis process for oil pipeline leakage is established based on logical reasoning:

(1) Leak identification

a. When the pressure  $p_1$  of annulus is greater than 100 psi (approximately 0.69 MPa), it is considered that there is a possibility of leakage, and further detection and analysis are required.

b. Test the annulus liquid level  $L$ , then perform pressure relief, and define the pressure relief gradient  $\Delta p$ . The liquid level test is performed every time the annular pressure drops by  $\Delta p$ , and the dynamic liquid level  $L$  in the annular space during the pressure relief process is obtained.

c. When the annulus pressure drops by a certain value, the pressure relief is stopped, and the annulus sound wave signal is detected. Extract the characteristic value of the acoustic signal (RMS, crest factor, form factor, peak-to-peak value, root square amplitude, etc.) and compare it with the unpressurized (No leakage condition) annulus acoustic signal to determine whether there is leakage downhole.

d. When a downhole leak is identified, perform helium tracer detection. If the helium is detected at the outlet of the tubing, there is a leak in the tubing; if helium is not detected at the outlet of the tubing, the production casing leaks, and the pressure in the annulus of the tubing is caused by the intrusion of the outer annulus fluid.

e. The pressure change in the outer casing is monitored during the pressure relief of the oil casing annulus. If the outer annulus pressure  $p_2$  changes, there is leakage in the production casing; on the contrary, if there is no significant change in  $p_2$ , there is no leakage in the production casing.

f. If there is no change in  $p_2$  and a downhole leak is identified or helium gas at the outlet of the tubing is detected, it means that only the tubing has a leak. If  $p_2$  changes significantly, and downhole leakage or helium return from the tubing outlet is identified, it indicates that both tubing and production casing are leaking. At this time, the production fluid will enter the oil casing annulus and technical casing annulus through the leak hole in turn., Causing the pressure of the outer annulus to rise sequentially. This situation poses the greatest threat to the safe production of offshore gas wells. Once this situation is discovered, it is necessary to immediately formulate strict control measures to prevent the pressure from further breaking through the outer annulus barrier and causing irreparable losses.

(2) Leak location and quantitative diagnosis process

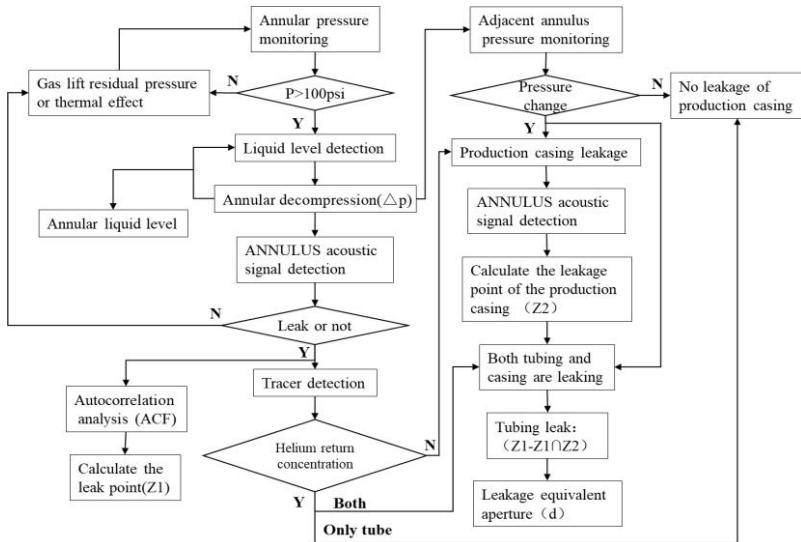
a. When it is recognized that there is a leaking sound wave in the annulus of the oil jacket, autocorrelation processing is performed on the leaking sound wave signal, the characteristic time of the leaking sound wave signal is extracted, and the annular sound velocity distribution curve is combined to estimate the oil casing leak position  $Z_1$ .

b. Similarly, when a leak is identified in the outer annulus, sound wave detection is performed to estimate the location of the leak point  $Z_2$ .

c. Further, the tracer test proves that when only the tubing leaks, the leaking position of the tubing is, using the liquid level  $L$  in the annulus of the oil casing, the leak point depth  $Z_1$ , and the tracer concentration  $C_t$  at the tubing outlet, combined with the wellbore and production Information, estimate the equivalent hole diameter  $d$  of the oil pipe leakage.

d. When it is recognized that the tubing and the production casing are leaking at the same time, the leakage point of the tubing at this time is  $(Z_1-Z_1 \cap Z_2)$ , that is, the leakage point on the production casing is eliminated first, and then the tubing leakage is quantitatively analyzed.

The diagnosis of the outer annulus follows the same process.



**Figure 2.** Comprehensive diagnosis process of oil casing leak

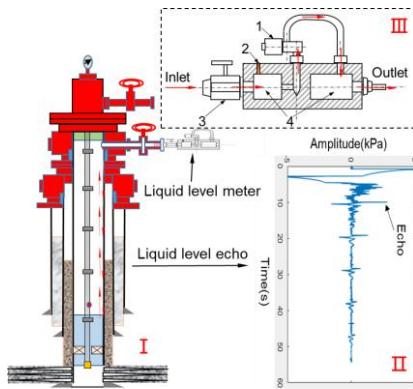
### 2.3. Liquid Level Detection

The principle of liquid level detection is based on the acoustic wave method, and the calculation formula is shown in formula (1). The key is to obtain the time  $t$  for the sound wave to propagate from the wellhead to the liquid surface, and the propagation speed  $v$  of the sound wave in the annulus. The speed of sound  $v$  can be calculated using equation (2) [6].

$$H = vt \quad (1)$$

$$v = \left[ \left( \frac{c_p}{c_v} \right) \left( \frac{RT_g}{M_g} \right) \left( Z_g + \rho \left( \frac{\partial Z_g}{\partial \rho} \right)_{T_g} \right) \right]^{1/2} \quad (2)$$

Where:  $c_p$  and  $c_v$  are the specific heat capacity at constant pressure and constant volume respectively,  $J/(mol \cdot K)$ ;  $R$  is the general gas constant  $J/(mol \cdot K)$ ;  $T_g$  is the temperature,  $K$ ;  $Z_g$  is the compressibility factor;  $\rho$  is the mole Density,  $mol/m^3$ ;  $M_g$  is molar mass,  $kg/mol$ , and the medium is annular air.



**Figure 3.** Principle of liquid level detection (1- Solenoid; 2- Acoustic transducer; 3- Ball valve; 4- Pressure chamber)

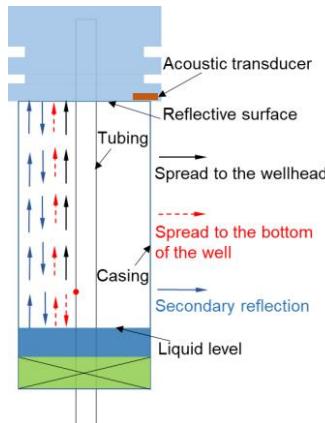
The sonic gun shown in Figure 3 (III) is used to launch a test wave into the annulus. When the test wave propagates to the bottom of the annulus, it will reflect on the liquid surface of the annulus. The liquid level echo signal is obtained at the wellhead of the annulus—Figure 3 (II). Determine the depth of the annulus liquid surface by extracting the echo time combined with the annulus sound velocity distribution law. Its working principle is: install it on the four-way gate valve of the Christmas tree tubing using a pipeline. After the ball valve is opened, there is a pressure difference between the two ends of the sonic gun. The high-pressure gas is released instantly by opening and closing the solenoid valve, and the high-pressure gas flows through the solenoid valve pressure fluctuations are generated during the channel and propagate into the annulus along the gas medium. The pressure fluctuations are the test waves in the liquid level test process.

#### 2.4. Principle of Leak Location

When the oil pipe leaks, high-pressure natural gas enters the production annulus through the leak hole. As shown in Figure 4, a continuous acoustic signal is generated at the leak hole. Part of the acoustic signal propagates along the annulus to the wellhead, and the other part propagates downhole to the annulus liquid surface, and is transmitted to the wellhead after being reflected by the liquid surface. Install a sound wave receiving sensor at the exit of the Christmas tree tubing pipe to obtain the sound wave signal in the annulus. The acoustic signal collected at the wellhead in the annulus is the superposition of the leaked acoustic wave after different delays. The delay has the following relationship with the relative position of the leak in the annulus [7]:

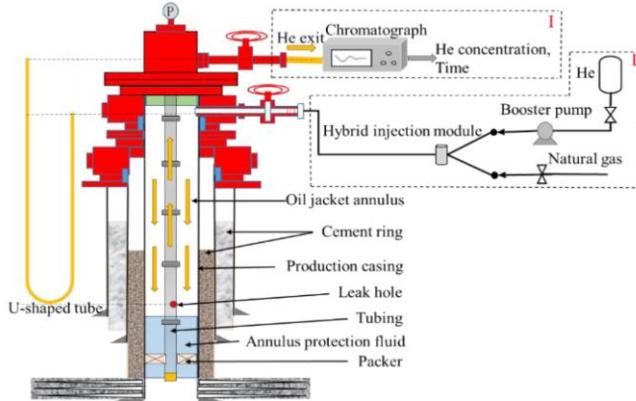
$$X = \bar{v}(t_2 - t_1)/2 \quad (3)$$

Where  $x$  is the distance between the leaking point and the reflecting surface of the tubing hanger;  $t_1$  is equal to the time of a return journey between the leaking sound wave from the leaking point to the annular liquid surface;  $t_2$  is equal to the one between the leaking sound wave hanging from the tubing to the annular liquid surface Return time.



**Figure 4.** Schematic diagram of oil pipe leakage location

## 2.5. Principles of Tracer Detection



**Figure 5.** Schematic diagram of tracer detection

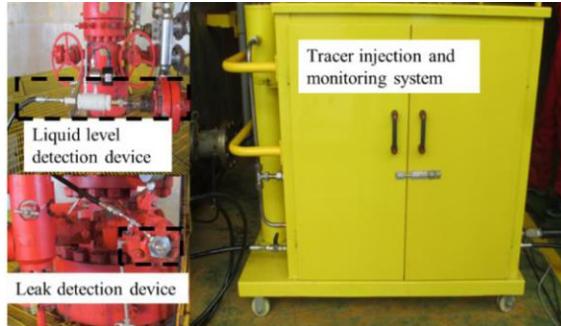
As shown in Figure 5, when the tubing leaks, a certain amount of tracer (helium is used as the tracer for leak detection in the text [8-10]) and natural gas mixture is continuously injected from the annulus wellhead. Driven by the air pressure, the mixed gas diffuses down the annulus to the leak point, passes through the leak point and enters the tubing, mixes with the natural gas in the tubing, and flows to the wellhead along with the produced gas. Collect and detect the concentration of helium in natural gas at the wellhead, and quantitatively analyze the leakage degree of the tubing according to production parameters, wellbore parameters and helium injection parameters. Assuming that the initial helium concentration is  $C_0$ , the injection time is  $t$ , the production parameter is  $Y$ , the wellbore parameter is  $B$ , the injection parameter is  $X$ , and the leakage is  $Q$ , the concentration of the returned helium can be expressed as:

$$C = C_0 f(t, B, X, Y, Q) \quad (4)$$

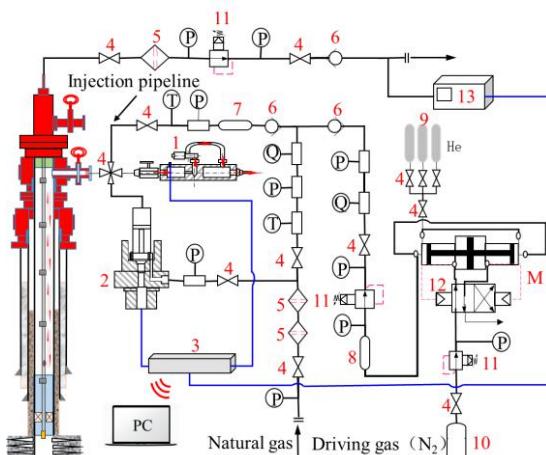
The reverse solution can be obtained Oil pipe leakage [11].

## 2.6. System

Integrating fluid monitoring, liquid level detection, leak detection and tracer detection and other parameters, the downhole tubing and casing leak diagnosis system developed is shown in Figure 6. Its working principle is shown in Figure 7, and the components are shown in Table 1. Show.



**Figure 6.** Physical map of downhole tubing and casing leak diagnosis system



**Figure 7.** Working principle diagram of downhole casing leakage diagnosis system

**Table 1.** Summary of components of the downhole tubing and casing leak monitoring and diagnosis system

No.	Name	No.	Name
1	Liquid level detection device	8	Gas Tank
2	Leak point detection device	9	Helium cylinder
3	surveillance system	10	Driving gas (N2)
4	Valve	11	Pressure reducing valve
5	Oil-gas separation device	12	Supercharger
6	Check valve	13	Chromatograph
7	Mixing tank	P	Pressure transmitter
T	Temperature transmitter	Q	Mass flowmeter

The detailed operation process is:

(1) Leak identification: Close the gate valve, drain the residual pressure on the outside of the gate valve, connect the liquid level detection device No. 1 shown in Figure

7 to the wellhead four-way annulus outlet, and open the gate valve after completing the tightness test of each pipeline. Quickly open and close the solenoid valve and collect the annulus sound wave signal. After the test is over, close the gate valve, vent the residual pressure in the external pipeline, and then remove the liquid level detection device.

(2) Leak sound wave detection: Close the gate valve, drain the residual pressure on the outside of the gate valve, install the threaded plug with acoustic wave sensor number 2 as shown in Figure 7 at the wellhead four-way annulus outlet, open the gate valve slowly to perform annulus acoustic wave test. In order to avoid the moment when the gate valve is opened, the high-pressure gas in the annulus impacts and destroys the diaphragm of the acoustic wave sensor, the annular air body is reversely introduced through the pressure relief hole on the threaded plug, so that the pressure in the four-way annulus outlet is equal to or It is close to the pressure in the wellbore annulus, reducing the possibility of damage to the acoustic wave sensor after being impacted.

(3) Tracer test: Switch the downhole tubing and casing leak diagnosis system to the tracer injection module. When the tracer is injected, a gas source well with a pressure higher than the measured annulus pressure is required to drive the tracer to rapidly diffuse down the annulus to the leak hole. Injecting the mixture of tracer and natural gas in a slug manner can greatly reduce the amount of tracer. As shown in Figure 7, the tracer 9 is mixed with the high-pressure natural gas in the mixing tank 7 in a certain proportion through the pressurizing mechanism 12. The booster pump is driven by compressed nitrogen 10. Open the gate valve and the injection pipeline valve, inject the tracer into the annulus, and record the injection time, injection flow rate, injection temperature and pressure and other fluid parameters. After a period of injection, the injection of the tracer is interrupted, and the high-pressure natural gas is continuously injected, which drives the tracer to move quickly into the annulus. While injecting the tracer, open the sampling valve connected to the outlet of the tubing, and use the gas chromatograph number 13 shown in Figure 7 to continuously monitor the natural gas produced at the wellhead. Since the gas chromatograph sampling tube is an ordinary low-pressure equipment, the produced natural gas needs to be filtered and depressurized, and the pressure of the sample after depressurization is about 0.1 MPa. Considering that the endothermic reaction of natural gas in the process of pressure reduction and flow restriction is likely to cause ice blockage in the pressure relief pipeline, the system adopts a needle valve-reducing valve combination multi-stage pressure reduction method. After the test is completed, close each valve in turn, and after evacuating the residual pressure in the system, remove the connected pipeline and restore the wellhead.

### 3. Engineering Practice and Results

Using the developed downhole tubing and casing leak monitoring and diagnosis system, 15 well-time detection experiments were carried out in the offshore annulus killing well. Taking the A5 detection result as an example, the test results are shown in Table 2, and the liquid level detection signal is shown in Figure 8, the time-domain diagram of the annulus acoustic signal and its autocorrelation curve are shown in Figure 9.

**Table 2.** Test results of Well A5

Well	Annulus	Pressure/MPa	Results		
			Liquid level /m	Leak location/m	Size/mm
A5	A	9.14	—	2070.71	853.3 2.5mm

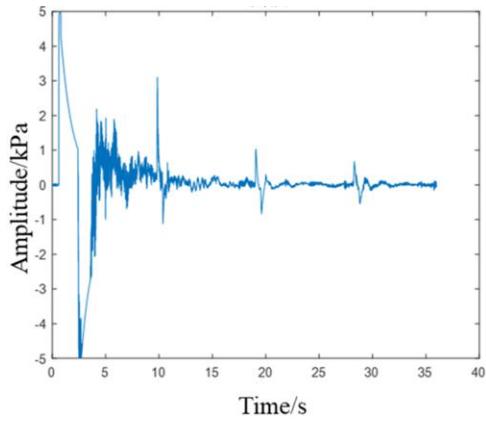


Figure 8. Annular liquid level test curve

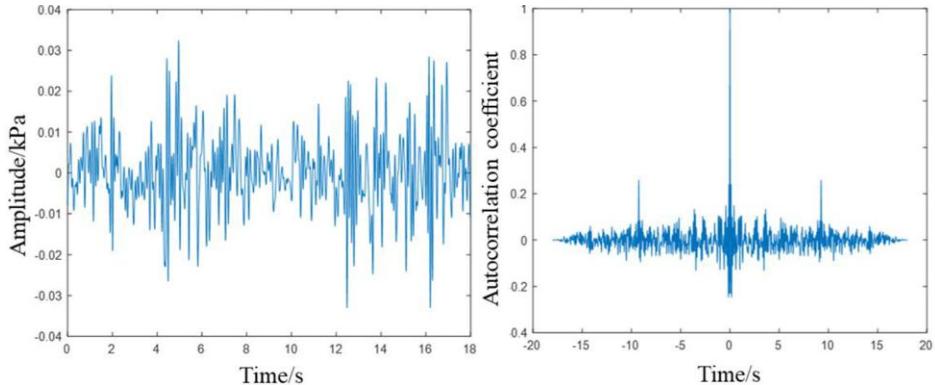
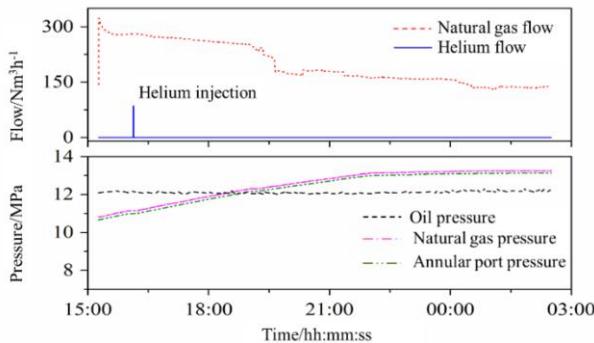
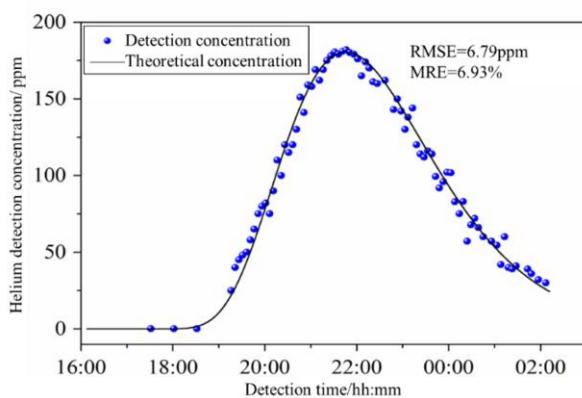


Figure 9. Time domain diagram of annulus acoustic signal and its autocorrelation curve

During the tracer test, continuously monitor the flow and pressure of the injected natural gas and the pressure in the wellhead annulus and tubing, and draw the change curve of each parameter as shown in Figure 10. A slug-type helium injection is used, the injection time is 137s, the slug length is 2.01m, and the flow of helium is zero after the injection is stopped. With the continuous injection of natural gas, the pressure in the annulus gradually increases, the injection flow rate gradually decreases, and the injection pressure and the annulus pressure are always consistent. With the continuous injection of natural gas, the injection pressure and the annulus pressure are slowly higher than the pressure in the tubing, and the pressures of the two stabilized at 13.13MPa. At this time, the pressure in the tubing was 12.22MPa. During this period, the gas chromatograph was used to continuously monitor the helium concentration in the tubing. Under this working condition, the peak return concentration of helium at the tubing outlet was about 177 ppm, and the peak time was 338 min. The helium detection concentration is summarized as shown in Figure 11. Based on the pressure, flow, temperature and other parameters monitored by the tracer, combined with the detection results of the liquid level and the depth of the leak, the equivalent aperture of the leak is calculated to be 2.5mm. Using this equivalent aperture to calculate the theoretical concentration curve of helium returning in the forward direction is shown in Figure 11, which is more consistent with the detection concentration curve of helium.



**Figure 10.** Variation curve of various parameters during the tracer test



**Figure 11.** Comparison of theoretical concentration and detection concentration of helium

#### 4. Conclusion

A single detection method cannot fully obtain downhole leakage information. The development of a gas well downhole tubing leakage monitoring and diagnosis system suitable for field use has always been the direction of field staff and scientific research workers.

(1) Integrating the existing fluid monitoring, liquid level detection, leak acoustic wave positioning and tracer detection technologies, a comprehensive diagnostic analysis process for tubing and casing leaks is established based on logical reasoning, in order to realize the downhole leak status (leak location, leak degree and the wellhead holographic diagnosis of the leak source) provides theoretical support.

(2) Integrated fluid monitoring, liquid level detection, leak acoustic wave positioning and tracer detection technologies, a downhole tubing and casing leak monitoring and diagnosis system was developed, and using this system, 15 detection experiments were carried out in the offshore annulus zone to kill wells. The data on the depth of the annulus liquid level, the depth of the downhole leak and the size of the leak are realized, and the holographic diagnosis of the downhole leakage state at the wellhead without moving the pipe string and affecting the production can be provided to ensure

the safe production of gas wells. Effective technical support provides an important basis for the next comprehensive treatment of well killing in the annulus zone.

## References

- [1] Hwang R J, Elsinger R J. Detecting production tubing leak by time resolved geochemical analysis of oils[C]. The Production Operations Symposium, Oklahoma, U.S.A., 1995: 67-79.
- [2] Bourgoyn A T, Scott S L, Regg J B. Sustained casing pressure in offshore producing Wells [C]. The 1999 Offshore Technology Conference, Houston, Texas, 1999.
- [3] API RP90, Annular casing pressure management for offshore wells[S]. Washington: API, 2006.
- [4] Zhu H, Lin Y, Zeng D, et al. Mechanism and prediction analysis of sustained casing pressure in "A" annulus of CO<sub>2</sub> injection well [J]. Journal of Petroleum Science & Engineering, 2012, 92-93:1-10.
- [5] Tarmoom I, Thabet H B, Samad S, et al. A comprehensive approach to well-integrity management in Adma-Opco[C]. The 15th SPE Middle East Oil & Gas Show and Conference, Manama, Bahrain, 2007.
- [6] ASSOCIATION A G. AGA Report No. 10 - Speed of sound in natural gas and other related hydrocarbon gases[R]. Transmission Measurement Committee, 2002: 12-17.
- [7] Liu D, Fan J C, Wu S N. Acoustic Wave-Based Method of Locating Tubing Leakage for Offshore Gas Wells. Energies, 2018, 11(12): 3454.
- [8] Kohl J, Newacheck R L, Anderson E E. Locating casing shoe leaks with radioactive argon[J]. Petroleum Transactions, AIME, 1995, 204: 213-216.
- [9] Shnaib F, Nadar M S, Sreekumar M P, et al. Successful application of CO<sub>2</sub> tracer technology for surveillance of gas lifted wells[C]. The SPE Production and Operations Conference and Exhibition, Tunis, Tunisia, 2010.
- [10] Li F. Research and Application of Gas Tracer Monitoring Technology[D]. Dongying: China University of Petroleum (East China), 2013.
- [11] Liu D. Resraech on Ground Detection and Diagnosis Method for Tubing and Casing Leakage in Offshore Gas Well[D], Beijing: China University of Petroleum (Beijing), 2019.

# Unreadable Segment Recognition of Single-Lead ECG Signals Based on XGBoost: Fusion of Shannon Energy Envelope and Empirical Mode Decomposition

Hanshuang XIE<sup>a,1</sup>, Jiayi YAN<sup>a</sup>, Huaiyu ZHU<sup>b</sup>, Qineng CAO<sup>a</sup>, Yamin LIU<sup>a</sup>, Yun PAN<sup>b</sup>  
and Fan WU<sup>a</sup>

<sup>a</sup>*Research and Development Department, Hangzhou Proton Technology, Co., Ltd.,  
Hangzhou, China*

<sup>b</sup>*College of Information Science and Electronic Engineering, Zhejiang University,  
Hangzhou, China*

**Abstract.** The quality of ECG signals is commonly affected by severe noise, especially for the single-lead ECG signals acquired from long-term wearable devices. Recognizing and ignoring these interfered signals can reduce the error rate of automatic ECG analysis system, and in addition, improve the efficiency of cardiologists. Based on XGBoost classifier, we propose an unreadable ECG segment recognition method using features extracted through Shannon Energy Envelope (SEE) and Empirical Mode Decomposition (EMD). An unreadable CarePatch™ ECG patch database is established, containing 8169 readable segments and 6114 unreadable segments with a length of 10 seconds. The XGBoost with 5-fold cross-validation is applied and obtained an accuracy of 99.51+/-0.15%. In conclusion, SSE and EMD features contribute to the unreadable segments recognition and alleviate the misdiagnosis of abnormal rhythms.

**Keywords.** Unreadable segment, single-lead ECG, wearable devices, XGBoost

## 1. Introduction

With the development of m-Health technology, wearable ECG monitoring devices of various physical forms such as card-type, patch-type, watch-type, etc. have appeared, which are widely applied to various clinical diagnosis scenes. By extending the monitoring time and scenarios, wearable ECG devices with limited lead channels compensate for diseases diagnose that are easily missed by resting 12-lead ECG and traditional 24-hour Holter, such as cryptogenic syncope, palpitation, paroxysmal atrial fibrillation, transient arrhythmia [1]. However, as daily activities are inevitable in long-term wearing, ECG signals collected by wearable devices are severely interfered by

<sup>1</sup> Corresponding Author, Hanshuang XIE, Research and Development Department, Hangzhou Proton Technology, Co., Ltd., Room 405, Building #8A, East Software Park, No. 90 Wensan Road, Xihu District, Hangzhou, China; E-mail: hanshuang.xie@protontek.com.

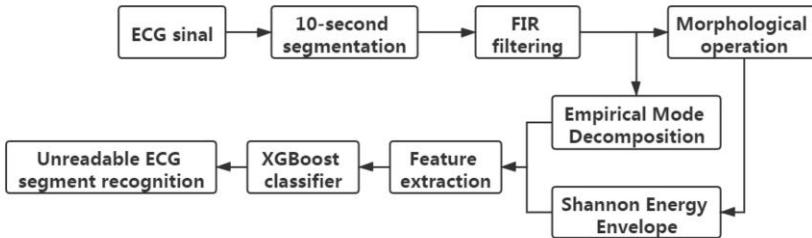
various noise, which will reduce the quality of signals and the work efficiency of cardiologists. Although shielding noisy segments can promote the performance of classification algorithm and reduce the false alarm rate [2], deleting such noise brutally may lead to false recognition of some diseases such as arrhythmia. Hence, to retain the effective signals as much as possible, only segments with R peaks completely overwhelmed by noise are regarded as unreadable segments.

Some early published works have proposed several generally applied ECG signal quality evaluation and noise estimation methods. Lee *et al.* [3] applied empirical mode decomposition (EMD) for extracting the first-order intrinsic mode function (F-IMF). However, the data they used is obtained from healthy persons, which cannot reflect the influence of noise on the misdiagnosis of arrhythmia. Li *et al.* [4] experimented with 13 signal quality indexes (SQIs) and their combinations from 12-lead ECG data based on support vector machine (SVM). Zhang *et al.* [5] adopted pSQI, kSQI and basSQI as indicators and confirmed that the single-lead ECG signal evaluation algorithm is capable of signal acquisition, denoising and QRS extraction. Zhao *et al.* [6] executed simple heuristic fusion to extract four SQIs and established the fuzzy vector for classification and assessment. Satija *et al.* [7] took autocorrelation function features into consideration. Moeyersons *et al.* [8] extracted three features from the autocorrelation function (ACF) and fed them into a RUSBoost classifier. The autocorrelation function features above can improve the assessment results effectively, but this time-consuming method presents challenges in long-term real-time ECG signal quality evaluation task. Zaunseder *et al.* [9] adopted 35 simple spectrum features, and obtained balance between accuracy and computational complexity with random forest (RF) classifier on 12-lead ECG data. They also stated that a more accurate classification is not feasible using their feature space. Zhang *et al.* [10] presented encoding Lempel-Ziv complexity (ELZC) method for feature extraction, combining other linear and nonlinear features and have proved that PE and ELZC features contribute most to noise recognition. Orphanidou *et al.* [11] utilized wavelet entropy features based on heart rate variability and SVM for signal quality classification, and verified the generalization of their model on different sing-lead ECG sensors.

In this paper, we present an unreadable segments recognition method of single-lead ECG signals based on XGBoost classifier, propose a new feature set extracted from Shannon Entropy Envelope, combined the features with that those extracted from time-domain and EMD. We also established an unreadable CarePatch™ ECG patch database using the data collected by the single-lead wearable patch devices. We conduct simulation experiments on the constructed dataset and achieve an accuracy of 99.51+/-0.15%.

## 2. Method

The proposed method as shown in Figure 1 contains the following steps: 1) divide ECG signals into 10 seconds' fragments; 2) filter signals with finite impulse response (FIR) band-pass filter with a cut-off frequency range between 0.67 Hz and 40 Hz; 3) for filtered signals SigFir, apply morphological opening and closing operations and thereafter, obtain Shannon Energy Envelope (SEE); On the other hand, apply EMD and obtain the first IMF component; 4) extract relevant features from SEE and EMD respectively; 5) train the XGBoost classifier; 6) detect the unreadable segments of ECG signals.



**Figure 1.** The block diagram of the proposed method.

## 2.1. Pre-processing

In his paper, we adopt FIR band-pass filter ( $0.67\text{Hz} \sim 40\text{Hz}$ ) as a basic filter to remove the noise [12]. It is confirmed that the FIR band-pass filter with a cut-off frequency of  $0.67\text{ Hz}$  can handle baseline wander well, and signals with a frequency lower than  $40\text{ Hz}$  retain most of the useful information [13].

## 2.2. Features

### 2.2.1. Feature Selection

By expanding earlier published works [3-6], we extract 28 signal quality indicators for unreadable segments recognition, as shown in Table 1.

**Table 1.** Features extracted from the proposed methods.

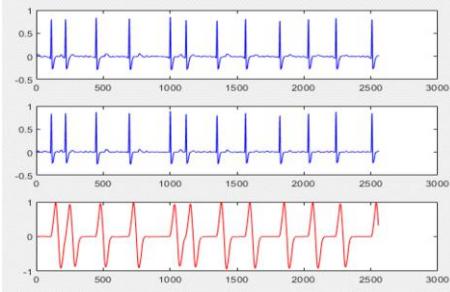
Methods	Feature Name	Dimension	Methods	Feature Name	Dimension
sigFir <sup>[4-6]</sup>	Kurtosis	1	SEE	meanAmp	1
	Skewness	1		stdAmp	1
	Entropy	1		ratioMeanStd	1
	validAmplitude	1		first8PeaksMeanStdRatio	1
	invalidRpeaksRatio	1		first8PeaksMean	1
EMD <sup>[3]</sup>	Mean of IMF1	3		first8PeaksStd	1
	Std of IMF1	3		first5PeaksMeanStdRatio	1
	Zero crossing rate	3		first5PeaksMean	1
	Shannon entropy	3		first5PeaksStd	1
				histCountRatio	1
				Large2mvRatio	1

### 2.2.2. Shannon Energy Envelope (SSE)

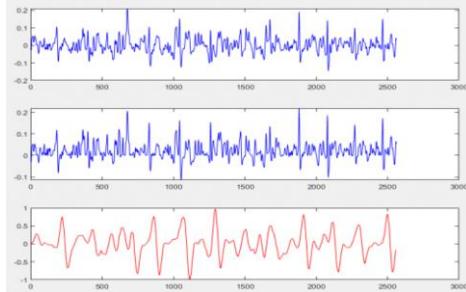
In this paper, Shannon Energy Envelope related features are adopted innovatively. The filtered signal SigFir is put through imopen and imclose operations to obtain sigMor, and then the Shannon envelope SigShannon is computed. Finally, 11 relevant features are extracted for the classification of unreadable segments. The steps are as follows:

- 1) Apply band-pass filter with a cut-off frequency of  $6\text{-}18\text{ Hz}$ , retain the energy of the QRS complex and obtain sigFilter;
- 2) Calculate the difference of the signal and normalize it to get sigDiffNor;
- 3) Compute Shannon entropy and obtain sigShannonEntropy;
- 4) Apply mean filtering with a window length of  $0.18\text{s}$  twice and obtain sigAvg;
- 5) Calculate the difference of sigAvg and normalize it, obtain the final Shannon envelope sigShannon;

6) Extract relevant features from sigShannon.



**Figure 2.** The sigShannon of readable segments (original, filtered and Shannon envelope)



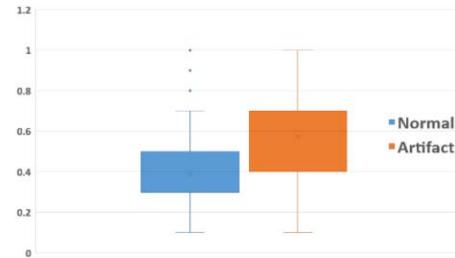
**Figure 3.** The sigShannon of unreadable segments (original, filtered and Shannon envelope)

Examples of the sigShannon of readable and unreadable segments are shown in Figure 2 and Figure 3.

We further extracted various features from the Shannon envelope sigShannon and finally selected 11 from them (as shown in Table 1) which have significant differences in the distribution of readable and unreadable segments respectively. The intuitively differences in feature distribution are shown in Figure 4 and Figure 5.



**Figure 4.** The distribution of first5PeaksMean feature on readable and unreadable segments



**Figure 5.** The distribution of histCountRatio feature on readable and unreadable segments

### 2.3. Modelling

In this paper, XGBoost is applied for classification, with boosting rounds n\_estimators = 101, learning rate = 0.5, maximum tree depth max\_depth=10, minimum sum of instance weight needed in a child min\_child\_weight = 8, nthread = 1, subsample = 0.85.

## 3. The Unreadable CarePatch™ ECG patch database

Due to the users' daily activities, single-lead ECG data collected by long-term wearable devices usually contain more forms of heartbeats and more complex noises, which poses a great challenge to automatic ECG signal analysis. However, it is difficult to make a performance comparison of unreadable segments recognition algorithms on single-lead ECG because of the lack of a long-term single-lead ECG database. Therefore, constructing a long-term unreadable segments of single-lead ECG database is of great significance for the algorithm to assist in unreadable segments recognition tasks.

In this study, we firstly construct a long-term single-lead ECG data set. Then we conduct the proposed unreadable segments recognition algorithm and the other recognition algorithms on the dataset for a comparison. This self-build data set is collected and established by NMPA-certified CarePatch™ patch devices (NMPA#ZJ20202070050), containing 14283 single-lead ECG data segments, with 8169 readable ones and 6114 unreadable ones respectively. Each segment is of 10 seconds long, and the distribution of the training set and test set is 12854 (5492 unreadable ones, 7362 readable ones) and 1429 (622 unreadable ones, 807 readable ones) respectively. This unreadable CarePatch™ ECG patch database is collected from 499 patients' single-lead ECG recordings from July, 2017 to April, 2020. These patients were requested to wear CarePatch™ patch devices for 7 days. The patch device collected non-standard lead ECG signals from up left of the chest with a 256 Hz sampling rate and 12-bit resolution [14].

## 4. Results

### 4.1. Results on Test Set

We divide the total data set into training set (90%) and test set (10%), and with 10-fold cross-validation on the training set, we obtained an average accuracy of 99.51 $\pm$ 0.15%. The generalization performance on the test set is presented with ROC curve depicted in Figure 6 (AUC=0.9999), the confusion matrix presented in Figure 7.

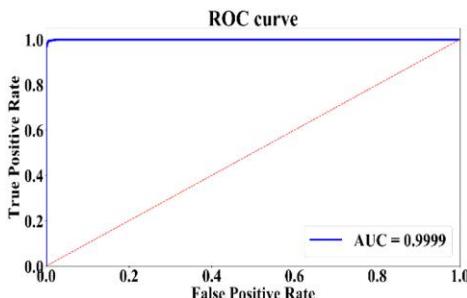


Figure 6. ROC curve on the test set.

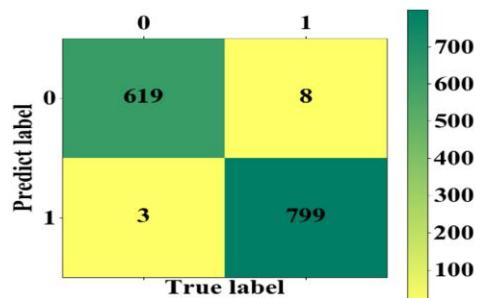


Figure 7. Confusion matrix on the test set

### 4.2. Results for Different Feature Combinations

To further identify the effectiveness of the proposed methods, we compared our features and the other feature set by combining SigFir features, EMD features and SEE features as illustrated in Table 2. The results are all obtained by XGBoost classifier, and we can see that the feature combination adopted by this paper contributes to the best performance on the unreadable segment recognition task.

Table 2. The performance of XGBoost on different feature combinations.

Feature set	Precision%	Recall%	F1
<b>SigFir</b>	97.27	97.27	0.973
<b>EMD</b>	98.74	98.74	0.987
<b>Shannon envelope</b>	97.62	97.62	0.976
<b>SigFir+EMD</b>	98.88	98.88	0.989

<b>SigFir+Shannon envelope</b>	99.09	99.09	0.991
<b>EMD+Shannon envelope</b>	98.96	98.95	0.990
<b>This paper</b>	99.23	99.23	0.992

## 5. Conclusion

The contributes of this paper can be summarized as followings:

- (1) Proposed an unreadable segment recognition of single-lead ECG signals based on XGBoost with the fusion of Shannon energy envelope features, empirical mode decomposition features and other time-domain features, which outperforms the state-of-the-art methods.
- (2) Constructed an unreadable CarePatch™ ECG patch database based on the medical-level single-lead ECG monitoring devices, which contains 7-day ECG signals and along with various interference and noise. The construction of such database is of significance to evaluate the unreadable segment recognition algorithm.

The training and testing code developed in this work is publicly available in the repository<sup>2</sup>.

## References

- [1] Chen XL, Liu M, Yang J. Clinical application of single lead long-term ambulatory electrocardiography. *J. Practical Electrocardiol.*, 2018; 27(04): 269-272+278.
- [2] Abdelazez M, Quesnel PX, Chan AD, Yang H. Signal quality analysis of ambulatory electrocardiograms to gate false myocardial ischemia alarms. *IEEE Trans. Biomed. Eng.* 2016 Aug 25; 64(6):1318-25.
- [3] Lee J, McManus DD, Merchant S, Chon KH. Automatic motion and noise artifact detection in Holter ECG data using empirical mode decomposition and statistical approaches. *IEEE Trans. Biomed. Eng.* 2011 Nov 10;59(6):1499-506.
- [4] Li Q, Rajagopalan C, Clifford GD. A machine learning approach to multi-level ECG signal quality classification. *Comput. Meth. Programs Biomed.* 2014 Dec 1;117(3):435-47.
- [5] Zhang Y, Zhao Z. Evaluation of single-lead ECG signal quality with different states of motion. In 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI) 2017 Oct 14 (pp. 1-7). IEEE.
- [6] Zhao Z, Zhang Y. SQI quality evaluation mechanism of single-lead ECG signal based on simple heuristic fusion and fuzzy comprehensive evaluation. *Front. Physiol.* 2018 Jun 14; 9:727.
- [7] Satija U, Ramkumar B, Manikandan MS. An automated ECG signal quality assessment method for unsupervised diagnostic systems. *Biocybern. Biomed. Eng.* 2018 Jan 1;38(1):54-70.
- [8] Moeyersons J, Smets E, Morales J, Villa A, De Raedt W, Testelmans D, Buyse B, Van Hoof C, Willems R, Van Huffel S, Varon C. Artefact detection and quality assessment of ambulatory ECG signals. *Comput. Meth. Programs Biomed.* 2019 Dec 1; 182: 105050.
- [9] Zaunseder S, Huhle R, Malberg H. CinC challenge—Assessing the usability of ECG by ensemble decision trees. In 2011 Computing in Cardiology 2011 Sep 18 (pp. 277-280). IEEE.
- [10] Zhang Y, Wei S, Zhang L, Liu C. Comparing the performance of random forest, SVM and their variants for ECG quality assessment combined with nonlinear features. *J. Med. Biol. Eng.* 2019 Jun;39(3):381-92.
- [11] Orphanidou C, Drobniak I. Quality assessment of ambulatory ECG using wavelet entropy of the HRV signal. *IEEE J. Biomed. Health Inform.* 2016 Oct 5;21(5):1216-23.
- [12] Venkatachalam KL, Herbranson JE, Asirvatham SJ. Signals and signal processing for the electrophysiologist: part I: electrogram acquisition. *Circ.-Arrhythmia Electrophysiol.* 2011 Dec;4(6):965-73.

<sup>2</sup> <https://gitee.com/xieHanshuang/unreadable-segment-recognition.git>

- [13] Romero FP, Romaguera LV, Vázquez-Seisdedos CR, Costa MG, Neto JE. Baseline wander removal methods for ECG signals: A comparative study. arXiv preprint arXiv:1807.11359. 2018 Jul 30.
- [14] Zhu H, Pan Y, Wu F, Huan R. Optimized electrode locations for wearable single-lead ECG monitoring devices: A case study using WFEES modules based on the LANS method. Sensors. 2019 Jan;19(20):4458.

# A Study on the Correlation Between Emotion and the Quality of Life of the Elderly Under the Artificial Intelligent Smile Recognition System: Protocol for a Descriptive Study

Ruitong CAI<sup>a</sup>, Ze ZHANG<sup>b,1</sup>, Lu MENG<sup>c</sup>, Kang GE<sup>c</sup>, Wenhao YIN<sup>a</sup>, Han QI<sup>a</sup>, Hongxin ZHAO<sup>a</sup> and Suning LI<sup>a</sup>

<sup>a</sup> Department of Geriatrics, Affiliated Hospital of Liaoning University of Traditional Chinese Medicine, No.33 Beiling Street, Huanggu District, Shenyang, 110033, China

<sup>b</sup> Graduate School, Liaoning University of Traditional Chinese Medicine, No.79 Chong Shandong Road, Huanggu District, Shenyang, 110032, China

<sup>c</sup> School of Information Science and Engineering, Northeastern University, Lane 11, Wenhua Road, Heping District, Shenyang, 110819, China

**Abstract.** With the aging of the population, the senior people's quality of life has become a hot issue in China and even in the world. This study, with emotion as the core, evaluated the senior citizens' quality of life in the form of scales to explore the relationship between emotion and human health. Artificial intelligent identification system has developed rapidly in this society and has made great contributions to the development of human society. Combining artificial intelligent identification system with medicine will contribute to the development of human public health, which is of great significance and value.

**Keywords.** Elderly, emotion, AI smile recognition system, quality of life

## 1. Introduction

With the aging of the population, the senior citizens' quality of life has become a hot issue of global concern. Quality of life refers to the individual's feeling experience of his own physical condition, psychological function, social ability and personal comprehensive condition in different culture and value system [1]. The physical and mental health status of the elderly, the relationship with the family and society are important factors affecting the senior citizens' quality of life population [2-4]. The

---

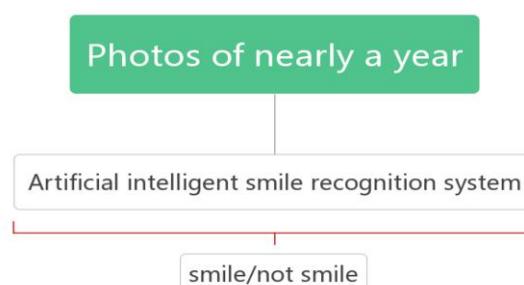
<sup>1</sup> Corresponding Author, Ze ZHANG, Affiliated Hospital of Liaoning University of Traditional Chinese Medicine,33 Beiling Street, Huanggu District, Shenyang City, Liaoning Province, China; E-mail: 13066796232@163.com.

elderly group is more likely to produce emotional loneliness than the young people, and the senior citizens' quality of life is often negatively affected by the gradual decline of various functions of the body and the loss of interpersonal relationships [5]. From many factors that may affect the senior citizens' quality of life, this study selected emotional factors as the core of the study, aiming to explore the impact of emotion on the senior citizens' quality of life.

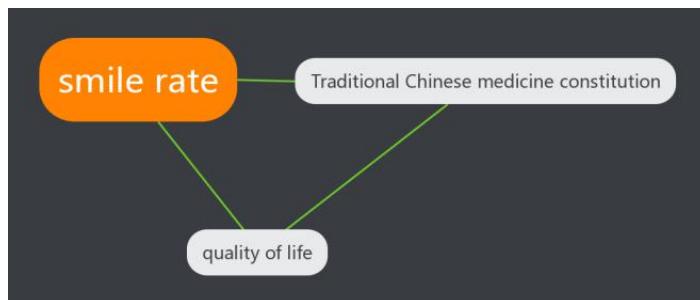
In order to explore the relationship between emotion and the life quality, we must judge the emotion of the participants. So we use the artificial intelligent smiling face recognition system to identify the photos of the participants. After comprehensive analysis, the participants will be judged as positive or negative personalities. In the early stage of the project, we cooperated with the Institute of Deep Learning of Northeastern University in the research of face recognition and facial expression classification, and completed the scientific research platform of facial expression intelligent recognition based on convolutional neural network. Next, the two sides will work together in the clinical, big database, methodology and other areas of further in-depth cooperative research, expected to conclude the correlation between subjects' emotions and physical health. Combining artificial intelligent identification system with medicine will contribute to the development of human public health, which is of great significance and value [6-8].

## 2. Methods

The objectives of this study are as follows: By using the AI smiling face recognition system to identify the photos of the elderly population for nearly a year. In recent years, there are studies about the scale and the application of various scales in medicine has become more and more mature, this study selected SF-36 scale to evaluate the senior citizens' quality of life [9]. SF-36 quality of life scale was developed by Boston health institute, USA. This study used the Chinese translation version of the social medicine teaching and research office of Zhejiang University medical school as a health evaluation tool for participants in this topic. The quality of life of each participant will be evaluated from 8 dimensions. And through the scale evaluation, the quality of life of the elderly will be evaluated so as to analyze the relationship between the quality of life and emotion of the elderly. The AI system workflow is shown in Figure 1. A three-aspect correlation study in Figure 2 was performed after the system calculated the percentage of smiling photos in the subjects' photographs.



**Figure 1.** AI system workflow chart.



**Figure 2.** Correlation study

### 3. Study Design

This is a multicenter, retrospective, descriptive study. The study was completed by Liaoning University of Traditional Chinese Medicine in cooperation with Northeastern University.

AI smiling face recognition process: the realization of AI smiling face recognition function is divided into the following three steps:

- Uniform picture size and format.
- Image capture of mouth and surrounding parts.
- Image recognition.

AI Smiling Face Recognition System Principle: We use a convolutional neural network (CNN) to build the model, which has eight layers, consisting of five convolution-pooling layers and three fully connected layers. CNN has three main components, including convolutional layer, pooling layer, and fully connected layer [10-11]. The convolutional network can adapt well to the small-scale translation of the image, that is, it has good translation invariance [12].

Original image is input into the CNN structure after regularization. Each neuron in each layer takes the output of the previous layer as input. CNN obtain feature maps through convolution operations, which can represent different image features. There are always more than one feature maps in the convolutional layers, thereby retaining the richer features of the image. After the convolutional layers, pooling layers are used to downsample image, which reduces the image resolution and the amount of parameters, moreover, it obtains the robustness of translation and deformation. As the growing of convolution layers and pooling layers, the number of feature maps increases, and the resolution decreases. Finally, the probability of different categories is output directly through the full connection layer, and the classification result is given at the output end.

Sampling design and setting: The sample size refers to the sample size estimation method of multivariate analysis, and the sample size required is generally 5-10 times the number of independent variables. In order to ensure the quality of the research, learn from the existing research, and consider the feasibility factors such as research cost and research cycle, a total of 200 samples were selected in this survey.

Inclusion criteria: (1) Older persons  $\geq 60$  years of age; (2) Plenty of photos for the past year; (3) Voluntary participation in trials and signature of informed consent for clinical investigations.

Exclusion criteria: (1) Plenty of photos for the past year, but most of the photos face occlusion or poor photo quality, AI the system cannot identify the person; (2) A person with a special face arising from a special illness; (3) Complicated with advanced malignant tumor, mental illness and severe neurosis.

Informed consent and basic information gathering: Before participating in the trial, the researcher will send the informed consent notification page and the signature page to the participants, which contains the contents, significance, procedure, risks and interests of the clinical study of the project. Fully understood, the participants and researchers will sign the confirmation on the signature page. Participants need to provide basic information to the researchers for later statistics and follow-up.

Assessment of participants' quality of life: Reference to expert opinions, combined with the reliability and validity evaluation of various quality of life scales, this experiment selected SF-36 quality of life scale to evaluate the quality of life of participants [13-15]. The scale assessed participants' quality of life from eight dimensions: bodily pain (BP), role physical(RP), physical functioning (PF), general health (GH), vitality (VI), mental health (MH), role emotional (RE), and social functioning(SF). The first four dimensions mainly reflect the degree of physical health (PCS), and the latter four dimensions mainly reflect the degree of mental health (MCS).

Completion of participants' quality of life scale: In order to ensure the accuracy of the assessment of participants' quality of life, the researchers were responsible for explaining the items in the scale and asking questions to the participants one by one. All the results were recorded by the researchers, and checked the items separately.

SF-36 scoring methods: First of all, each item is coded, the answer checked by the participants is scored, and the eight dimensions are respectively converted according to the score conversion formula. The higher the score after the conversion, the better the status of the participants in this dimension.

Data management: (1) The participant's previous photo information is voluntarily provided to the researcher by the participant himself. The researcher should ensure the privacy of the photo, after being identified by the AI smiling face recognition system, the original photo is archived, and the AI identification results are accurately and timely recorded and transferred to the data administrator. (2) Quality of life scale (SF-36) will be completed by clinical researchers. Clinical researchers should ensure that the data are accurately, completely and timely loaded into the quality of life scale (SF-36), while keeping the original records, and the data administrator is handed over after the completion of the above scale is reviewed by the clinical inspector; (3) To ensure that the data are true and valid, participants are required to specify the date of completion of the form and the specific time taken to complete the form.

Data statistics: Collate the questionnaire, use the SPASS23.0 software for data collation and analysis. The general data of participants will be analyzed by frequency and percentage. By means of mean, standard deviation and rate, the smiling photo proportion, emotional state and quality of life will be analyzed. A single-factor analysis of participants' emotions is with two independent samples t test or variance analysis. A partial correlation analysis will be used for the correlation between participants' emotions and their quality of life.

Quality control: Researchers must undergo uniform training, uniform recording and judgment criteria. All observations in clinical trials should be verified, and the conclusions in clinical trials are derived from the original data.

#### 4. Discussion

This is a retrospective study. By appropriate methods, our study will give a comprehensive answer to the association between emotion and senior citizens' quality of life. In China, and even in the world, the number of elderly people is increasing year by year, and the study of the elderly people's life quality should be more detailed [16]. The senior citizens' life quality includes, but not only includes, physical health status. This study not only explores the influence of emotion on the quality of life of the senior, but also explores the relationship between occupation, marital status, living conditions and the emotions of the elderly, thus providing a variety of ideas for improving the senior citizens' quality of life [17-18].

Our study has some limitations, and the 200 participants selected from different centers and regions are still the majority of participants in Northeast China. Besides, the AI intelligent smiling face recognition system used in this study is not ideal for image recognition with occlusion and side, and is not good for some images with small smile amplitude.

#### 5. Trial Status

The trial is ongoing and currently recruiting.

#### 6. Ethics

This trial is conducted in accordance with the Helsinki Declaration and relevant Chinese clinical trial research norms and regulations. The study was reviewed and approved by the Ethics Review Committee of Chinese Registered Clinical Trials (ChiECRCT20200010). Written informed consent of all participants will be obtained before data collection.

#### References

- [1] Panzini RG, Mosqueiro BP, Zimpel RR, Bandeira DR, Rocha NS, Fleck MP. Quality-of-life and spirituality. *Int Rev Psychiatry*. 2017; 29(3): 263-282. doi:10.1080/09540261.2017.1285553.
- [2] Vagetti GC, Barbosa Filho VC, Moreira NB, Oliveira Vd, Mazzardo O, Campos Wd. Association between physical activity and quality of life in the elderly: a systematic review, 2000-2012. *Braz J Psychiatry*. 2014;36(1):76-88. doi:10.1590/1516-4446-2012-0895.
- [3] Boggatz T. Quality of life in old age - a concept analysis. *Int J Older People Nurs*. 2016; 11(1):55-69. doi:10.1111/opn.12089.
- [4] Gómez-Morales A, Miranda JMA, Pergola-Marconato AM, Mansano-Schlosser TC, Mendes FRP, Torres GV. Influencia de las actividades en la calidad de vida de los ancianos: revisión sistemática [The influence of activities on the quality of life of the elderly: a systematic review]. *Cien Saude Colet*. 2019;24(1):189-202. doi:10.1590/1413-81232018241.05452017.
- [5] Salomoni F, Addelyan Rasi H, Hosseinzadeh S. Empowering elderly Iranians through a social group work intervention: A trial study to assess the effect of the intervention on participants' quality of life. *Health Soc Care Community*. 2018;26(6):917-924. doi:10.1111/hsc.12617.
- [6] Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism*. 2017; 69S: S36-S40. doi:10.1016/j.metabol.2017.01.011.
- [7] Wang F, Preininger A. AI in health: state of the art, challenges, and future directions. *Yearb Med Inform*. 2019; 28(1):16-26. doi:10.1055/s-0039-1677908.

- [8] Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol.* 2017 Jun 21; 2(4): 230-243.
- [9] Doosti-Irani A, Nedjat S, Nedjat S, Cheraghi P, Cheraghi Z. Quality of life in Iranian elderly population using the SF-36 questionnaire: systematic review and meta-analysis. *East Mediterr Health J.* 2019 Jan 23; 24(11): 1088-1097. doi:10.26719/2018.24.11.1088.
- [10] Ilyas N, Shahzad A, Kim K. Convolutional-neural network-based image crowd counting: review, categorization, analysis, and performance evaluation. *Sensors (Basel).* 2019 Dec 19; 20(1): 43. doi:10.3390/s20010043.
- [11] Liang M, Zhou T, Zhang F, Yang J, Xia Y. *J Biomedical Eng.* 2018;35(6):977-985. doi:10.7507/1001-5515.201710060.
- [12] Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK. Medical image analysis using convolutional neural networks: a review. *J Med Syst.* 2018 Oct 8.; 42(11):226. doi:10.1007/s10916-018-1088-1.
- [13] Matcham F, Scott IC, Rayner L, et al. The impact of rheumatoid arthritis on quality-of-life assessed using the SF-36: a systematic review and meta-analysis. *Semin Arthritis Rheum.* 2014; 44(2): 123-130. doi: 10.1016/j.semarthrit.2014.05.001.
- [14] Campolina AG, Ciconelli RM. O SF-36 e o desenvolvimento de novas medidas de avaliação de qualidade de vida [SF-36 and the development of new assessment tools for quality of life]. *Acta Reumatol Port.* 2008;33(2):127-133.
- [15] Alesii A, Mazzarella F, Mastrilli E, Fini M. The elderly and quality of life: current theories and measurements. *G Ital Med Lav Ergon.* 2006;28(3 Suppl 2):99-103.
- [16] Netuveli G, Blane D. Quality of life in older ages. *Br Med Bull.* 2008;85:113-126. doi:10.1093/bmb/ldn003.
- [17] Gurland B. The impact of depression on quality of life of the elderly. *Clin Geriatr Med.* 1992;8(2):377-386.
- [18] Farquhar M. Elderly people's definitions of quality of life. *Soc Sci Med.* 1995;41(10):1439-1446. doi:10.1016/0277-9536(95)00117.

## Part 3.

# Modern Management based on Big Data II – Late Submissions

This page intentionally left blank

# Spatio-Temporal Analysis of CO<sub>2</sub> Emission Driving Force in Various Provinces in China Using the Extended STIRPAT-GWR Model

Yulin ZHANG<sup>1</sup>

*Business School, Guilin University of Electronic Technology, Guilin 541004, China*

**Abstract.** To fill the shortcomings of traditional research that ignores the driver's own spatial characteristics and provide a theoretical support to formulate suitable emission reduction policies in different regions across China. In this pursuit, based on the panel data of provincial CO<sub>2</sub> emission in 2007, 2012, and 2017, the present study employed the extended environmental impact assessment model (STIRPAT-GWR model) to study the effect of population, energy intensity, energy structure, urbanization and industrial structure on the CO<sub>2</sub> emissions in 29 provinces across China. The empirical results show that the effect of drivers on the CO<sub>2</sub> emissions exhibited significant variations among the different provinces. The effect of population in the southwest region was significantly lower than that of the central and eastern regions. Provinces with stronger energy intensity effects were concentrated in the central and western regions. The effect of energy structure in the eastern and northern regions was relatively strong, and gradually weakened towards the southeast region. The areas with high urbanization effect were concentrated in the central and the eastern regions. Furthermore, significant changes were observed in the high-effect regions of the industrial structure in 2017. The high-effect area showed a migration from the northwest and northeast regions in 2007 and 2012, respectively, to the southwest and southeast regions in 2017. Urbanization showed the strongest effect on the CO<sub>2</sub> emissions, followed by population and energy intensity, and the weakest effect was exhibited by the energy and industrial structure. Thus, the effects of population and energy structure showed a downward trend, in contrary to the effect of urbanization on the CO<sub>2</sub> emissions in China.

**Keywords.** CO<sub>2</sub> emissions, STIRPAT-GWR model, spatial analysis drivers, China

## 1. Introduction

China has proposed the goal of reaching its peak CO<sub>2</sub> emissions by 2030. However, China's accelerating industrialization, urbanization and growing population have continuously increased the total energy consumption. Fossil energy consumption generates huge amounts of CO<sub>2</sub>.

Many scholars have done a lot of relevant research on the CO<sub>2</sub> emissions, and they have also achieved fruitful results. The driving factors of the CO<sub>2</sub> emissions have been

---

<sup>1</sup> Corresponding Author, Yulin ZHANG, Guilin University of Electronic Technology, Guilin 541004, China; Email: 1502904627@qq.com . Innovation Project of GUET Graduate Education: 2020YCXS071.

studied based on the factor decomposition methods and environmental impact assessment models [1], [2], [3], which mainly included population, energy, industry, urbanization, and economy. Many scholars have also studied the spatial differences of the CO<sub>2</sub> emissions based on the traditional regression models [4], and studied the spatial autocorrelation of CO<sub>2</sub> emissions based on the Moran index [5]. The results of the study indicate that there were significant spatial differences and autocorrelation in the CO<sub>2</sub> emissions. Despite the fruitful results, the existing research still faces certain shortcomings as follows: (1) Traditional models have shortcomings in space research. When it comes to space research, most scholars conduct analysis and research based on the traditional Ordinary Least Square model, but the OLS model does not involve the geographic location of the region. Traditional estimation models could have possibly presented biased results or even unreliable considering the spatial analysis[6]. (2) The traditional research ignored the spatial characteristics of the drivers themselves. Studies have shown that in spatial analysis, geographically weighted model was better than the traditional OLS regression [7]; and the environmental impact assessment model (STIRPAT) is not only relatively simple, but also can determine the impact of multiple human activities on the environment [8]. Therefore, the present study combined the traditional environmental impact assessment model with the geographically weighted model along with the spatial autocorrelation to build an integrated model, in order to study the spatial and temporal characteristics of the impact of the driving factors on CO<sub>2</sub> emissions, which overcomes the shortcomings of the traditional models in spatial research.

## 2. Data

The data of total population, urban population, gross domestic product, gross output value of secondary industry, gross output value of tertiary industry in this study were obtained from the China Statistical Yearbook. The data on coal, oil, and natural gas consumption was obtained from the China Energy Statistical Yearbook. The data of CO<sub>2</sub> was calculated from the three primary energy consumption using the method recommended by IPCC 2006. Since the data of Taiwan Province, Hong Kong Special Administrative Region, Macau Special Administrative Region, and Tibet National Autonomous Region were difficult to obtain, they were not included in the data set in this study. There were no provinces around Hainan that share the same border and points. In the conceptualization of spatial relationship, adjacent elements need to meet the characteristics of common edge and common point. Therefore, the data of Hainan Province was discarded.

## 3. Research methods

Our data shows that there was a serious multicollinearity between the urbanization and per capita GDP. Considering that urbanization can reflect more problems, and the model fits well after removing per capita GDP, we mainly studied the demographic factors: total Population (PEO, expressed by the total population of the province); Economic factor: Urbanization (URB, expressed by the ratio of urban population to the total population of the province); Technical factor: Industrial structure (INST, expressed by the total output value of the secondary industry and the third The proportion of industry); the impact of

energy intensity (ENIN, total energy consumption compared with the regional GDP) and energy structure (ENST, the proportion of coal consumption to total consumption) on CO<sub>2</sub> emissions.

The parameter estimation model can be expressed as:

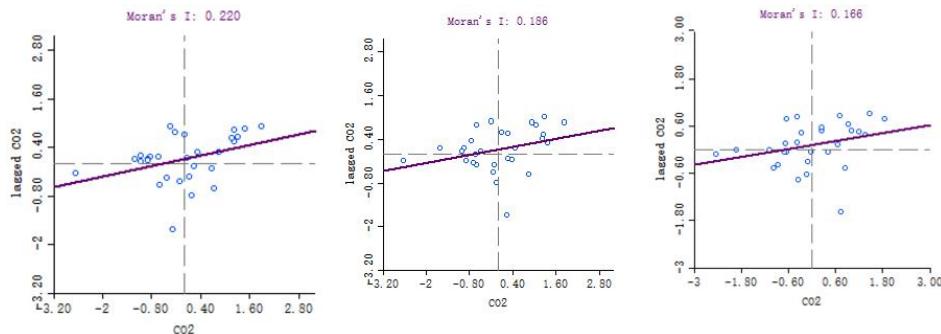
$$\ln C = \ln \alpha_0(u_i, v_i) + \alpha(u_i, v_i) \ln(x_i(PEO)) + \beta(u_i, v_i) \ln(x_i(URB)) \\ + \gamma_1(u_i, v_i) \ln(x_i(INST)) + \gamma_2(u_i, v_i) \ln(x_i(ENIN)) \\ + \gamma_3(u_i, v_i) \ln(x_i(ENST)) + \ln \mu_i$$

In this study, we proposed a new integrated model that combines the traditional OLS model and the GWR model to explore the direction and extent of the impact of social factors such as population, economy and energy on the CO<sub>2</sub> emissions.

## 4. Empirical results

### 4.1 Spatial correlation analysis

In 2007, 2012, and 2017, the Moran Index was 0.220, 0.186, and 0.166, respectively, which were all positive numbers, indicating that there was a positive correlation between the CO<sub>2</sub> emissions of the provinces, and geographic weighted regression models were available. However, the Moran index gradually declined, indicating that the degree of spatial autocorrelation of CO<sub>2</sub> emissions between the provinces was gradually weakening.



**Figure 1.** Moran index in 2007, 2012 and 2017

### 4.2 Detection of variable multicollinearity

The GWR model requires no multicollinearity between the variables. The multicollinearity of test is shown in Table 1. The VIF value can reflect, whether there is multicollinearity between variables. When the VIF value is greater than 7.5, it indicates that there are redundant variables. The results of the multicollinearity test of the variables showed in Table 1. The VIF value of each variable in 2007, 2012, and 2017 was less than 7.5, indicating that there was no problem of multicollinearity among the variables in the model.

**Table 1** Test results of multicollinearity of variables

Variable	2007	2012	2017
Population	1.92	2.13	1.78
Energy intensity	2.22	2.57	2.24
Energy structure	1.65	2.50	3.35
Urbanization	1.89	1.98	1.78
Industrial structure	1.38	1.71	2.47

#### 4.3 Results

The summary statistics of the STIRPAT-GWR coefficients show that the adjusted R<sup>2</sup> values in 2007, 2012, and 2017 were all greater than 0.95, indicating that the integrated model exhibited a good degree of fit. In addition, there were significant differences between the minimum and maximum values of the explanatory variables among the 29 provinces in each year, further indicating that it is necessary to consider the geographic location information.

**Table 2** Summary of the statistics of model coefficients

Variable	2007		2012		2017	
	Mean	Max-Min	Mean	Max-Min	Mean	Max-Min
Population	1.162320	0.305388	1.142193	0.386277	1.084614	0.332365
Energy intensity	0.857523	0.210014	0.883835	0.313516	0.849800	0.283530
Energy structure	0.578695	2.335581	0.320238	0.805641	0.247680	0.849750
Urbanization	1.941086	2.030032	2.054936	1.501717	2.377061	1.762383
Industrial structure	0.270392	0.238898	0.282393	0.577279	0.218559	0.770032
Adj R <sup>2</sup>	0.958723		0.952656		0.957677	

We visualized the regression coefficient of the model in ArcGIS, and the results are shown in Fig 5. According to the natural discontinuity method, the 29 provinces were divided into three categories, namely, the high impact factor area (dark), medium impact factor area (medium), and low impact factor area (light). The red map, blue map and purple map show the results of 2007, 2012 and 2017 respectively. A blank area in the map indicated that there was no data in the area. The results of the drivers are as follows: The impact of population on the CO<sub>2</sub> emissions showed a positive correlation trend (Liu et al. 2021), and the influence of the southwest region was significantly lower than that of the central and eastern regions. In 2007, high-impact areas were concentrated in the central and northeastern regions, and in 2012 and 2017, it was reduced to six central provinces. Energy intensity exhibited a positive impact on the CO<sub>2</sub> emissions, the average impact was about 0.8, and the high impact areas were concentrated in the central and western regions. The energy structure showed a positive effect on the CO<sub>2</sub> emissions. High-impact areas were concentrated in the eastern and northern regions, while the influence of the southeast was gradually decreasing. Urbanization showed a positive driving effect on the CO<sub>2</sub> emissions, and the intensity of the impact was very strong. High-impact areas were concentrated in the central and eastern regions. It can be seen that the high-impact areas of the industrial structure have undergone a major change in 2017. The high-impact areas have moved from the northwest and northeast regions in 2007 and 2012, respectively, to the southwest and southeast regions in 2017. The final quantitative change caused a qualitative change, and the industrial structure changed significantly during the 13<sup>th</sup> Five-Year Plan.

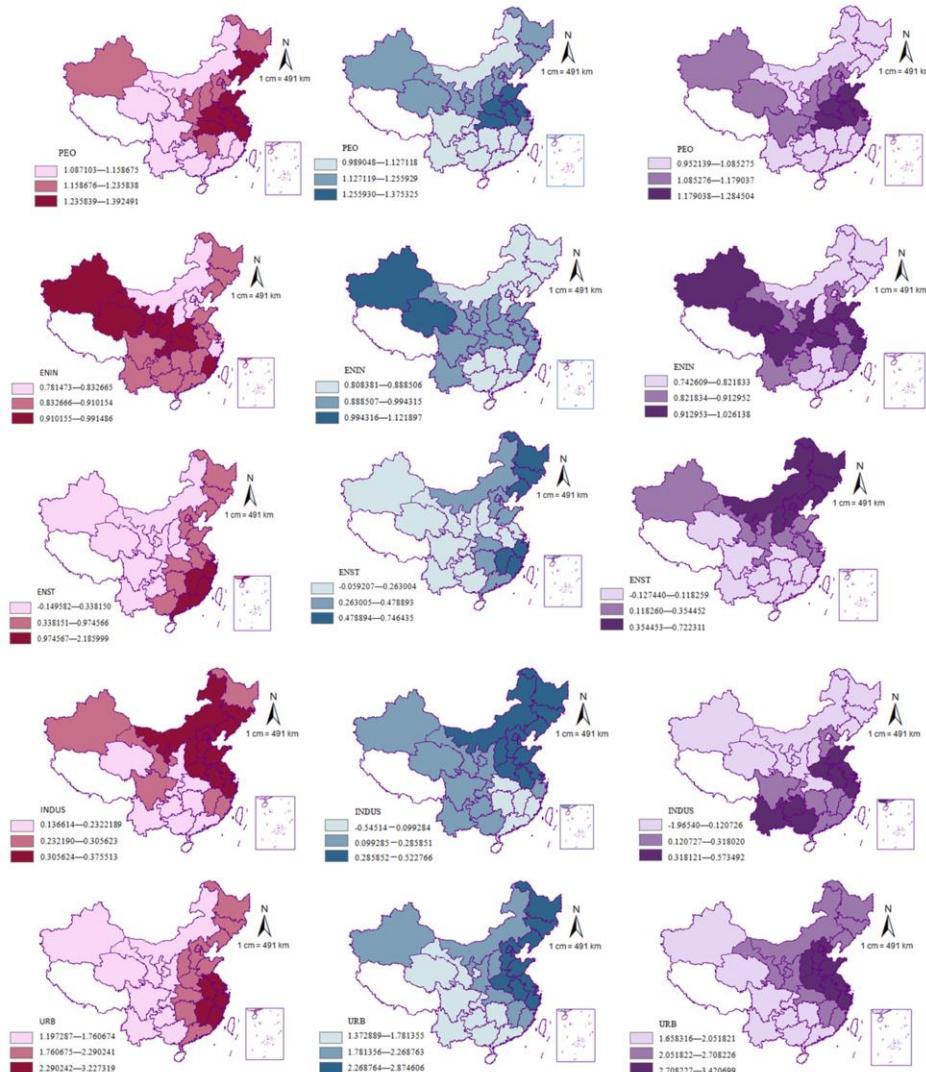


Figure 2. STIRPAT-GWR results in 2007, 2012 and 2017

## 5. Conclusions

In the present study, we employed the extended environmental impact assessment model to study the effect of population, energy intensity, energy structure, urbanization and industrial structure on the CO<sub>2</sub> emissions in 29 provinces across China. Our empirical results showed that: (1) There was a positive correlation and high concentration between the CO<sub>2</sub> emissions in various provinces across China. (2) Among the five factors, urbanization exhibited a highest impact on the CO<sub>2</sub> emissions, followed by population and energy intensity, while energy and industrial structure showed the weakest impact. (3) The five drivers had obvious spatial differences and regional concentration

characteristics on the intensity of CO<sub>2</sub> emissions. (5) The influence showed an upward or downward trend with time. Therefore, in areas where population, energy intensity, energy structure, urbanization and industrial structure influence gather, multi-provincial linkage will achieve good emission reduction effects. In the process of urbanization, more environmentally friendly and green building materials and more advanced energy-saving technology products are introduced will help urban pollution control. Increase the ratio of tertiary industry will improve energy efficiency.

## Reference

- [1] Luo YL, Zeng WL, Hu XB, Yang H, Shao L. Coupling the driving forces of urban CO<sub>2</sub> emissions in Shanghai with logarithmic mean Divisia index method and Granger causality inference. *Journal of Cleaner Production*.2021 Mar; 298:126843.
- [2] Patiño LI, Alcántara V, Padilla E. Driving forces of CO<sub>2</sub> emissions and energy intensity in Colombia. *Energy Policy*.2021 Feb; 151:112130.
- [3] Johannes L, Axel S, Andreas B. The role of demographic and economic drivers on the environment in traditional and standardized STIRPAT analysis. *Ecological Economics*.2020 Aug; 178:106811.
- [4] Hu MJ, Li RZ, You WH, Liu YB, Lee CC. Spatiotemporal evolution of decoupling and driving forces of CO<sub>2</sub> emissions on economic growth along the Belt and Road. *Journal of Cleaner Production*.2020 Aug; 277:123272.
- [5] Wang C, Wood J, Wang YJ, Geng XR, Long XG. CO<sub>2</sub> emission in transportation sector across 51 countries along the Belt and Road from 2000 to 2014. *Journal of Cleaner Production*. 2020 May; 266:122000.
- [6] Liu QW, Wu SM, Lei YL, Li ST, Li L. Exploring spatial characteristics of city-level CO<sub>2</sub> emissions in China and their drivers from global and local perspectives. *Science of the Total Environment*. 2021 Feb;754:142206.
- [7] Xu B, Lin B. Investigating spatial variability of CO<sub>2</sub> emissions in heavy industry: Evidence from a geographically weighted regression model. *Energy Policy*. 2020 Nov;149: 112011
- [8] Yang LX, Xia H, Zhang XL, Yuan SF. What matters for carbon emissions in regional sectors? A China study of extended STIRPAT model. *Journal of Cleaner Production*,2018 Jan;180(1):595-602..

# Investigating the Impact of Carbon Subsidy Policy on the Decision-Making of Remanufacturing Supply Chain

Yu ZHANG<sup>a,1</sup>, Tianshan MA<sup>a</sup> and Syed Abdul Rehman KHAN<sup>b</sup>

<sup>a</sup>*School of Economics and Management, Chang'an University, Xi'an, China*

<sup>b</sup>*School of Management, Xuzhou University of Technology, Xuzhou, China*

**Abstract.** This research is to investigate the decision making of the members of remanufacturing supply chain under the government involvement. Different scenarios are analyzed in this research, and it is found that the subsidy for carbon emission reduction can increase the WPs (waste products) reusing. When the recycler participates in remanufacturing supply chain, the cost of remanufacturer will be shared and through centralizing the decision making, the carbon emission reduction will be enhanced and the whole supply chain's profit will decrease. So it is suggested that the government need to adjust the subsidy for carbon emission reduction in terms of the quality level of WPs and the cooperation between recycler and remanufacturer is suggested, especially in the high-value waste remanufacturing supply chain .

**Keywords.** Remanufacturing, carbon emission reduction, supply chain, subsidy

## 1. Introduction

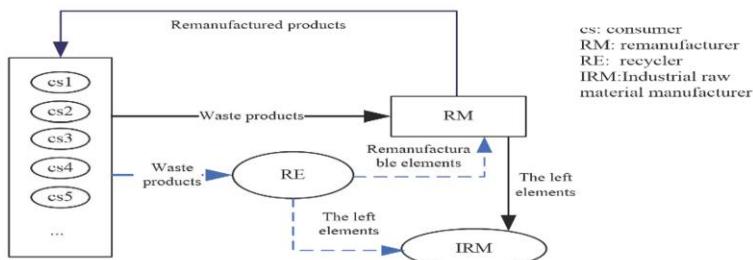
The recovery processes of WPs (waste products) are mainly remanufacturing of the remanufacturable elements and crushing and reusing the left elements. Remanufacturing is to recover the remanufacturable elements' quality and performance through repair processes and technologies. While crushing and reusing the left elements is to improve material purity through material separation technology and use in industrial raw materials manufacturing. Both methods of processing the WPs are friendly to the sustainability of environment and economy [1]. Though the WPs reusing are significant for the environmental and economic situation, the cost in WPs processing is an influential factor for the implementation of WPs reusing industry [2]. The processes of WPs reusing as raw materials includes WPs acquisition, classification, crushing, and material separation, which is a long procedure and costly [3]. The WPs reusing industry needs enough investment and more entities to share the cost and risk of WPs reusing industry. With environmental problems constantly emerging, the government starts to play an indispensable role in the development of WPs reusing industry. More and more governments start to provide subsidy for the carbon emission reduction in industrial production process, including WPs reusing processes [4-5]. They encourage the WPs to be reused through the instrument of policy. For instance, the original equipment manufacturer have to acquire and reuse their wasted products.

<sup>1</sup> Corresponding Author, Yu Zhang, School of Economics and Management, Chang'an University, China; Email: 984730042@qq.com

In fact, Zhou et al [6] concluded that the market competitiveness of the independent remanufacturer when the OEM (original equipment manufacturer) transfers the authority of remanufacturing to independent remanufacturer for a fee. This research focuses on the decision making of WPs RMSC (remanufacturing supply chain) considering the government involvement. Through analyzing the different practical scenarios of RMSC operation modes, this intend to find out the way to improve the efficiency and effectiveness of WPs reusing in terms of value creation and carbon emission reduction. This research takes an insight into the remanufacturing industry and provides an practical perspective to understand the different participants' decision making in RMSC. This paper is organized as below. In the second section, the methodology and the analysis for the decision making in RMSC is given. In the third section, the discussion and conclusion are provided.

## 2. Methodology

In WPs reusing industry, the quality level of WP is uniformly distributed in (0, 1), which is based on the researchers' data from several recycling companies. The recovered waste commodities are disassembled and classified into remanufacturable parts and the rest. The quality level of remanufactured elements is consistent with that of the recycled WPs. Remanufactured elements will eventually become marketable remanufactured products through the remanufacturing processes, and the left elements will be purchased by industrial raw material manufacturers. As production raw materials, both channels will produce carbon emissions, but the production process of producing new products with the materials from WPs will be less compared with that with developed primary materials. Moreover, the carbon emission reduction capacity of higher-quality remanufacturable element remanufacturing is better than the left element reusing in manufacturing industry. Further in remanufacturing, it is positively correlated with the quality level of remanufacturable elements. The government will give financial subsidies according to the carbon emission reduction in the production process. Here, the RMSC with different participants involvement is shown in figure 1.



**Figure 1.** The RMSC with different participants involvement.

Compared with new products produced from primary materials, carbon emissions from remanufacturing and crushing and reuse will be less. For remanufacturing, with the decrease of quality level, the time of remanufacturing will be longer and the process will be more complex, which causes more energy consumption.

## 2.1. Parameters setting

$Q$ , the volume of WPs in the market.

$P$ , the price of remanufactured product

$\delta$ , the quality level of a waste product, which is uniformly distributed in  $(0, 1)$ .

$q_{rm}$ , the number of WPs acquired by remanufacturer.

$q_{rc}$ , the number of WPs acquired by recycler.

$q_m$ , the number of remanufacturable elements that are selected by remanufacturer.

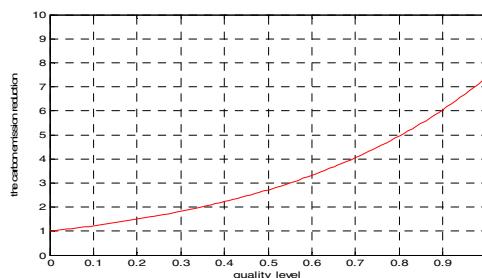
$q_r$ , the number of remanufacturable elements that remanufacturer purchases for remanufacturing.

$p_w$ , the constant parameter of acquisition price of WPs from consumers.

$p_r$ , constant parameter of purchasing price for remanufacturable element when the remanufacturer buys remanufacturable elements from recycler.

$p_o$ , the purchasing price of the left elements for industrial raw material manufacturer buying the left elements of WPs.

$e$ , the carbon emission reduction for each participants in RMSC. In figure 2, the relationship between quality level  $\delta$  with the carbon emission reduction in remanufacturing is given. The processing of each left element' carbon emission reduction from the WP will be the same.



**Figure 2.** the relationship between quality level  $\delta$  with the carbon emission reduction in remanufacturing.

$c_{rm}$ , constant parameter for remanufacturing cost when after testing, a remanufacturable element has no need to repair. And the remanufacturing cost of a remanufacturable element is  $\frac{1}{\delta^4} c_{rm}$   $c_p^m$ , the processing cost for recycled WPs when the remanufacturer does the WPs acquisition.

$c_p^r$ , the processing cost for recycled WPs when the recycler does the WPs acquisition.

$s_0$ , the constant parameter of the subsidy for carbon emission reduction in one remanufacturable element remanufacturing. The subsidy for WP processing is  $s_0$ , and the subsidy for one remanufacturable element remanufacturing is  $s_0 e^{2\delta}$ .

### 2.2.1. Scenario 1

When the remanufacturer operates the RMSC. As shown in figure 1, the WPs will be acquired and processed by the remanufacturer. Then the remanufacturable elements will be remanufactured and the left elements will be sold to the IRM (industrial raw material manufacturer). so, the profit function for the remanufacturer is,

$$\begin{aligned}\pi_{rm}^1 &= q_m(P + s_o \int_{\delta_o}^1 e^{2\delta} d\delta - c_{rm} \int_{\delta_o}^1 \frac{1}{\delta^4} d\delta) + (p_o + s_o)(2q_{rm} - q_m) \\ &- q_{rm}(c_p^m + \int_{\delta_{rm}}^1 p_w d\delta)\end{aligned}\quad (1)$$

Because the quality level  $\delta$  is uniformly distributed in (0, 1),  $q_{rm} = (1 - \delta_{rm})Q$ ,  $q_m = (1 - \delta_o)q_{rm}$ ,  $1 > \delta_{rm}$ ,  $\delta_o > 0$ . To find the optimal decision making on WP acquisitions and remanufacturing quantity, the partial derivatives are calculated.

$$\begin{aligned}\frac{\partial \pi_{rm}^1}{\partial \delta_o} &= (1 - \delta_{rm})Q[(p_o + s_o) - (P + s_o \int_{\delta_o}^1 e^{2\delta} d\delta) + c_{rm} \int_{\delta_o}^1 \frac{1}{\delta^4} d\delta + \\ &(1 - \delta_o)(c_{rm} \frac{1}{\delta_o^4} - s_o e^{2\delta_o})]\end{aligned}\quad (2)$$

$$\begin{aligned}\frac{\partial \pi_{rm}^1}{\partial \delta_{rm}} &= (1 - \delta_{rm})p_w Q - Q[(1 - \delta_o)(P + s_o \int_{\delta_o}^1 e^{2\delta} d\delta - c_{rm} \int_{\delta_o}^1 \frac{1}{\delta^4} d\delta) + \\ &(p_o + s_o)(1 + \delta_o) - (c_p^m + \int_{\delta_{rm}}^1 p_w d\delta)]\end{aligned}\quad (3)$$

When  $\frac{\partial \pi_{rm}^1}{\partial \delta_o} = 0$ ,  $\frac{\partial \pi_{rm}^1}{\partial \delta_{rm}} = 0$ , the optimal decision on the share of remanufacturing elements  $\delta_o$  and the share of WP acquisitions  $\delta_{re}$  can be obtained. Then  $q_{rm}$ ,  $q_m$  and the maximized profit  $\pi_{rm}^1$  can be calculated.

### 2.2.2. Scenario 2

When the recycler participates in waste RMSC. As shown in figure 1, the recycler does the WPs acquisition and processing, and then the remanufacturer purchases the remanufacturable elements from the recycler, and the left elements are sold to IRM.

When the two cooperators choose to do decentralized decision making in RMSC,  $q_m = q_r$  and the profit functions for the recycler and remanufacturer respectively are,

$$\pi_{re}^2 = q_r \int_{\delta_o}^1 p_r d\delta + (p_o + s_o)(2q_{re} - q_r) - q_{re} \int_{\delta_{re}}^1 p_w d\delta - q_{re} c_p^r \quad (4)$$

$$\pi_{rm}^2 = q_m(P + s_o \int_{\delta_o}^1 e^{2\delta} d\delta - \int_{\delta_o}^1 p_r d\delta - \int_{\delta_o}^1 c_{rm} \frac{1}{\delta^4} d\delta) \quad (5)$$

To find the optimal decision making on their decision variables, the derivatives for the two profit functions (2) and (3) are used to get the Nash Equilibrium between the two parties.

$$\begin{aligned}\frac{d\pi_{re}^2}{d\delta_{re}} &= (1 - \delta_{re})p_w Q - Q[(1 - \delta_o) \int_{\delta_o}^1 p_r d\delta + (p_o + s_o)(1 + \delta_o) - \\ &(c_p^m + \int_{\delta_{rm}}^1 p_w d\delta)]\end{aligned}\quad (6)$$

$$\begin{aligned}\frac{d\pi_{rm}^2}{d\delta_o} &= -(1 - \delta_{rm})Q[P + s_o \int_{\delta_o}^1 e^{2\delta} d\delta - \int_{\delta_o}^1 p_r d\delta - \int_{\delta_o}^1 c_{rm} \frac{1}{\delta^4} d\delta] + \\ &Q(1 - \delta_{rm})(1 - \delta_o)[p_r + c_{rm} \frac{1}{\delta_o^4} - s_o e^{2\delta_o}]\end{aligned}\quad (7)$$

So, when  $(1 - \delta_o)(2p_r + c_{rm} \frac{1}{\delta_o^4} - s_o e^{2\delta_o}) - P - \frac{1}{2}s_o(e^2 - e^{2\delta_o}) + \int_{\delta_o}^1 c_{rm} \frac{1}{\delta^4} d\delta = 0$ , the optimal decision on the share of remanufacturing elements  $\delta_o$  from recycler's waste can be obtained. Then, for the recycler, the share of WP acquisitions  $\delta_{re}$  from the total WPs can be obtained, when  $2p_w(1 - \delta_{re}) - (1 - \delta_o)^2 p_r - (p_o + s_o)(1 + \delta_o) + c_p^m = 0$ . Similarly,  $q_{re}$ ,  $q_m$  and the maximized profit  $\pi_{rm}^2$ ,  $\pi_{re}^2$  in decentralized mode can be calculated.

When the two cooperators choose to do centralized decision making in RMSC, the profit functions for whole RMSC respectively are,

$$\begin{aligned}\pi_c &= q_m(P + \int_{\delta_o}^1 e^{2\delta} d\delta - \int_{\delta_o}^1 c_{rm} \frac{1}{\delta^4} d\delta) + (p_o + s_o)(2q_{re} - q_m) \\ &\quad - q_{re}(\int_{\delta_{re}}^1 p_w d\delta + c_p^r)\end{aligned}\quad (8)$$

No matter in scenario 1 or 2, for the whole RMSC, the carbon emission reduction is function (5).  $e_o$  is the constant parameters for carbon emission reduction.

$$e_s = q_m e_o \int_{\delta_o}^1 e^{2\delta} d\delta + e_o(2q_{re} - q_m) \quad (9)$$

In centralized mode of scenario 2, it can be seen that the profit function (8) has only one difference with function (1). So the results can be obtained from the analysis result of scenario 1. In addition, when the  $\pi_c$  is maximized, the Pareto optimal profit for the RMSC might be achieved.

### 3. Discussion and conclusion

This research analyzed the profit creation and carbon emission reduction of RMSC in different scenarios. When the remanufacturer operates RMSC, he will decides the numbers of WPs he will acquire from consumers and the number of remanufacturable elements will be selected to be remanufactured. Compared with analysis result of the centralized decision making mode in scenario 2, in scenario 1, the remanufacturer can achieve the maximized profit of  $\pi_c$  through improving the WPs processing capability to decrease the cost  $c_p^m$ . In fact, the participation of recycler is to share the cost burden of remanufacturer and provide his advantage of WPs processing to RMSC, especially for the relatively expensive WPs, such as waste cars.

While, for the recycler, the profit is from the remanufacturer's purchasing of remanufacturable elements and IRM's purchasing of left elements but he needs to undertake the cost of all WPs acquisition and processing of whole RMSC. When the recycler consider his own profit, the WP acquisitions will be less [7]. While for the remanufacturer, though he is no longer needs to take the WPs acquisition, but the profit is just from the remanufactured products selling. The remanufacturing quantity is very difficult to increase when the remanufacturer just consider his own profit when making decision. Thus the carbon emission might decrease. When the taken the whole RMSC's profit into consideration, the analysis will be the same as scenario 1, and the results of remanufacturing quantity and the WP acquisitions will increase [6-7]. So when the decision making can be centralized the cost burden sharing and carbon emission reduction both can be achieved. While it is needed to consider the profit

sharing between the two parties (remanufacturer and recycler). The subsidy from the government for the carbon emission reduction is also the profit source for the RMSC operators, which can increase the number of WPs acquisition from consumers and remanufacturing quantity. Though, constrained by the quality level of WPs, the carbon emission is mainly affected by the number of WP acquisition and remanufacturing quantity.

So, it is suggested that the remanufacturer with insufficient capital can choose to cooperate the recycler to operate the RMSC. The information sharing between remanufacturer and recycler is also advocated. When the remanufacturer has sufficient capital to operate the RMSC, the processing capability should be enhanced to decrease the cost, such as the waste transportation, dissembling and classification. The governments is suggested to adjust its subsidy for the carbon emission reduction according to the quality level of WPs of the market in order to balance the financial expenditure and environment improvement.

## Reference

- [1] Dou G, Cao K. A joint analysis of environmental and economic performances of closed-loop supply chains under carbon tax regulation. *Comput & Industrial Eng.* 2020; 146: 1-13.
- [2] Zhu M, Li X, Ma J, Xu T, Zhu L. Study on complex dynamics for the waste electrical and electronic equipment recycling activities oligarchs closed-loop supply chain. *Env Sci and Pollut Res.* 2021; 1-21.
- [3] Masanta M, Giri BC. A manufacturing–remanufacturing supply chain model with learning and forgetting in inspection under consignment stock agreement. *Oper Res.* 2021; 1-25.
- [4] Li B, Geng Y, Xia X., Qiao D. Wang H. Comparatively Analyzing the Impact of Government Subsidy and Carbon Tax Policy on Authorized Remanufacturing. *Int J Env Res and Pub Health.* 2021; 18(16): 1-18.
- [5] Shekarian E., Marandi A., Majava J. Dual-channel remanufacturing closed-loop supply chains under carbon footprint and collection competition. *Sus Prod and Consump.* 2021; 28: 1050-1075.
- [6] Zhou Q, Meng C, Yuen K F. Remanufacturing authorization strategy for competition among OEM, authorized remanufacturer, and unauthorized remanufacturer. *Int J Prod Econ.* 2021; 242: 1-16.
- [7] Yan G, Ni Y, Yang X. Pricing and recovery in a dual-channel closed-loop supply chain under uncertain environment. *Soft Comput.* 2021; 25(21): 13679–13694.

## Subject Index

acquisition method	16	classification	583
adaptive graphs	655	clause	496
adaptive network	557	closed itemset	667
adaptive network model	596	CO <sub>2</sub> emissions	703
additive manufacturing	126	college students	590
affordability	416	compensation coefficient	442
AHP	435	compensation standard	442
AI smile recognition system	695	complex network	605
air pollution	317	complexity	496
all zero coefficient blocks	221	computational electromagnetics	57
analysis	360	computational tool	112
Analytic Hierarchy Process (AHP)	104	conjunctive normal form (CNF)	496
anisotropic diffusion	583	construction analytics	274
anomaly detection	96	construction delays	241
antonyms	502	construction management	
application performance	427	program	241
artificial neural network	241, 266	construction project proficiency	266
association rules	667	consumer behavior	248
audio watermarking	527	contrasting map	502
autonomous driving	539	control of adaptation	557
autonomous vehicle	539	convolutional neural network	49
behavior decision-making	40	cooperative game	393
Beidou satellite	256	correlation	632
belt and road initiative	169	cosine similarity	488
benefit sharing coefficient	458	cost analysis	368
BERT model	385	COVID-19	133, 162
bibliometric studies	119	CPT	527
big data	57, 256, 667	CS	527
big data analytics	169	customer loyalty	248
blockchain	133, 520	customer perceived value	248
boundary element method	57	customer satisfaction	248
brain-gut axis	596	cutting L1-norm	645
Brazil's navy	104	cybersecurity capacity	144
Building Information Modeling (BIM)	234	cybersecurity maturity models and interactive management	144
calculation	360	data analytics	188
carbon emission	360	data mining	667
carbon emission reduction	709	daytime	539
cascade dams	471	debt conservatism	32
catering	248	decision making	126
causal reasoning	481	decision support systems	406
central air conditioning	622	decrease mechanism	605
China	169, 703	deep learning	40, 385, 583

deep neural networks	569	goodman contact element	282
deep reinforcement learning	40	government subsidies	393
deformation transfer coefficient	282	grading method	427
diagnosis system	676	green building	360
dimensionality reduction	645, 655	hash function	520
discriminant analysis	645, 655	health certificate	133
disease diagnosis	481	HEVC	221
document image binarization	569	higher education	590
domain classifier	569	holographic diagnosis	676
double circulation pattern	66	hydrogen expressway	368
e-payment	205	hydrogen refueling stations	368
earthquake resisting design of pipe	282	hydrogen supply chain	368
ecological compensation mechanism game model	442	hydrological monitoring system	256
economic growth	169, 351	hyperledger	133
elbow	282	ice storage	622
elderly	695	impact	317
ELECTRE-MOr	126	inbound tourism	317
elliptic curve digital signature	520	income levels	205
emergency	458	industrial economic cooperation	66
emotion	695	information processing routes	3
energy consumption	351	information retrieval	24
energy storage model	622	integrated energy planning	375
entropy weight method	458	intelligent transport system	539
entropy weight TOPSIS method	341	interaction of pipe and soil	282
entry limitation	195	interdisciplinary framework	234
equity financing	178	Internet evolution	228
Ethereum	133	Internet+	256
evaluation of water resource	458	Internet-based	228
experience for reference	416	investments	162
feed formulation	614	isotropic diffusion	583
FFT	527	knowledge economy	16
final price arbitration method	442	knowledge graph	481
financial flexibility	32	knowledge reasoning	481
financing efficiency	178	KPI dashboard	274
flood	256, 471	Krylov subspace methods	57
flood hazard	435	leadership	228
flood-prone areas	435	lean construction principles	234
free trade agreement	77	liability identification	305
frequent itemset	667	linear programming	614
fuzzy comprehensive evaluation	458	literal	496
fuzzy logic	119	location model	368
Galois lattice	667	LWT	527
gas well	676	M&A performance	178
generation Z	205	machine learning	96, 154, 329
GF satellite	256	mainframe	96
GIS	435	management competency model	266
global financial crisis	32	management system	427
		manifold learning	502
		mass housing developments	241

mathematical programming	112	redundant resources	16
megacities	317	reinforcement learning	40
MELCHIOR	154	remanufacturing	709
mental model	557	renewable	375
modern ranking models	24	revolution of organization	228
monitoring	676	risk assessment	471
Multi-criteria Decision Aiding (MCDA)	126	risk management	188
Multi-criteria Decision Analysis (MCDA)	104, 119, 154	rural potable water treatment plant	427
multicriteria decision aid	89, 162	sample reconstruction	645, 655
narrative analysis	511	SATD	221
network evolution	605	satisfiability (SAT) problem	496
neural network	274	SAW	435
night time	539	search and analytics	24
non-compensatory	89	search engine	385
non-cooperative game	393	self-modeling network	557, 596
object detection	539	sensitivity analysis	266, 274
omni-channel behaviour	3	sentiment analysis	24
online car-hailing service	195	set pair analysis (SPA)	471
operation period	360	shaking table test	282
operational tunnel structure safety	329	shapley value	393
optimization	375	sharing bikes	296
ordinal ranking methods	89	sharing economy	195
outranking	154	shock wave	213
p-median	112	showrooming	3
pandemic	162	silk road	169
panel data simultaneous equations model	351	single-lead ECG	688
Parkinson's disease	596	smart contract	133
particle swarm optimization	614	smart finance	40
personality	590	social and economic value	458
personnel selection	119	soil CEC	632
perturbation analysis	488	spatial analysis drivers	703
plot dynamics	511	spectral clustering	488
plot extraction	511	state-owned enterprises	228
plot structure	511	statistic information	632
preconditioning	57	STIRPAT-GWR model	703
predicting model	632	streamflow forecast	406
price regulation	195	stress management	590
pricing mechanism	416	Structural Equation Modeling (SEM)	234
project performance	274	subsidy	709
project water price	416	SUDE model	341
public transportation stations	296	supply chain	709
QIM	527	support system	96
QR decomposition	527	surface integral equations	57
quality of life	695	sustainability	341
RDO quantization	221	tacit knowledge	16
real-time	539	tariff relief	77
		tax risk prediction	49
		taxi service	195

technology adoption	205	UTAUT	205
the supply chain of water diversion		value quantity	40
project in undeveloped areas	393	variable	496
the Yangtze River	256	variable fuzzy sets (VFS)	471
the Yangtze River Economic Belt	256	video integrity	520
timing financing theory	178	warship	104
trade complementarity	77	water pollution	305
trade complementarity index	66	water price elasticity	416
traffic congestion	213	water quality model	305
traffic flow	213	water resource dispatch	406
trans-boundary river	305	water resources	341
travel characteristics	296	wearable devices	688
triple constraints of project		webrooming	3
management	234	word embeddings	502
tubing leakage	676	word representations	502
tunnel engineering	329	word segmentation	385
tunnel structure disease prediction	329	workmanship defects	241
uniform quantization	221	XGBoost	688
unreadable segment	688	Z-RAYS	96
unsupervised domain adaptation	569		

## Author Index

Adina, E.M.	266	Fahmi, A.	527
Apeh, E.	144	Fan, J.	676
Bahrani, M.	24	Feng, Y.	406
Bai, X.	282	Fernandez, J.	274
Basílio, M.P.	89, 112, 119, 126, 154, 162	Fu, Q.	393, 442
Bhalwankar, R.	557	Garrido-Munoz, C.	569
Bissuel, C.	375	Ge, K.	695
Budiman, G.	527	Gerka, A.	96
Cai, R.	695	Goetzinger, C.	596
Calvo-Zaragoza, J.	569	Gomes, C.F.S.	89, 104, 112, 119, 126, 154, 162
Cao, Q.	688	Gou, J.	481
Cao, Z.H.	632	Gu, Y.	317
Carpentieri, B.	57	Guo, J.	77, 375
Castellanos, F.J.	569	Ham, S.	583
Chen, C.	40	Han, F.	368
Chen, G.	488	Han, S.	583
Chen, J.	406	Hao, S.	305
Chen, S.	427	He, B.	632
Chen, X.	282	He, T.	481
Chen, Y.	435, 471	He, Z.H.	632
Chen, Z.	385	Hong, B.-W.	583
Choi, H.-T.	583	Hong, Y.	406
Chudnovsky, B.	548	Huang, D.	282
Costa, A.P. de A.	162	Irawati, I.D.	527
Costa, I.P. de A.	89, 104, 112, 119, 126, 154, 162	Jaworski, D.	96
Courtot, F.	375	Ji, Y.	368
Cui, W.	32	Jiang, M.	66, 77, 317
de Almeida, I.D.P.	162	Jiang, W.	195
De Jesus, K.L.M.	234, 241, 266, 274	Jiao, B.	329
de Mattos Bento Soares, L.	112	Juan, W.	169
Ding, D.	590	Kalathinathan, A.	169
Dogan, H.	144	Karamitsos, I.	133
Dong, L.	341	Khan, S.A.R.	709
dos Santos, M.	89, 104, 112, 119, 126, 154, 162	Kim, H.	583
Drumond, P.	126	Kowal, B.	96
Du, X.	329	Kuraś, P.	96
Duan, K.	645, 655	Lai, W.	66
Duan, Z.	221	Lawrence, L.	520
Enciso, A.I.	234	Lee, H.	583
Estrada, J.	539	Lee, N.	583
		Leopold, G.	96
		Lewicz, M.	96

Li, A.	248	Qiu, Y.	605
Li, H.	385, 614	Raju, V.	169
Li, J.	385, 393, 442	Ramachandran, KK.	169
Li, K.	458	Rodrigues, M.V.G.	104, 119, 154
Li, L.J.	632	Roelleke, T.	24
Li, Q.	248	Samenko, I.	502
Li, S.	228, 695	Sánchez-Hernández, A.	569
Li, W.J.	632	Seck, D.A.N.	667
Liao, Y.H.	632	Shan, N.	221
Lim, R.U.	241	Shang, Q.	178
Liu, A.	228	Shao, G.	645, 655
Liu, D.	676	Sheng, W.	406
Liu, J.	16	Shi, B.	329
Liu, S.	393	Shreelekshmi, R.	520
Liu, X.	481, 590	Shrestha, S.	169
Liu, Y.	688	Silva, D.L.	234, 241, 266, 274
Lu, C.	427	Somasundaram, M.	188
Luo, J.	393, 442	Strzałka, D.	96
Luo, N.	49	Su, H.	329
Lyu, T.	605	Sudha, D.	188
Ma, B.-y.	622	Sun, Y.	442
Ma, F.	676	Tak, J.	583
Ma, T.	709	Tallón-Ballesteros, A.J.	v
Ma, X.	228	Tan, Y.	228
Maeda, S.M. do N.	89, 119, 154, 162	Tang, A.	282
Malysheva, A.	511	Tikhonov, A.	502, 511
Masykuroh, K.	527	Tong, X.	416
Mecija, A.N.	234	Toniolo, K.	596
Mei, J.	256	Treur, J.	557, 596
Mendoza, J.O.	234	Tripathi, A.	539
Meng, L.	695	Tsuboi, T.	213
Menn, N.	548	Umamaheswari, S.L.	188
Mohamed Junaid, K.A.	188	Uvarov, S.I.	496
Moreira, M.Â.L.89, 112, 119, 154,	162	van Ments, L.	557
Narhutaradhol, P.	205	Viejo-Fernández, N.	3
Naseir, M.A.B.	144	Villaverde, B.S.	234, 266
No, J.	583	Wang, D.	368
Noumi Sandji, N.S.	667	Wang, G.	360
Opina Jr, G.	539	Wang, H.	481
Ou, W.	645, 655	Wang, L.	329
Pan, Y.	688	Wang, R.	605
Papadaki, M.	133	Wang, X.	296
Peng, C.	645, 655	Wang, Y.	458
Pereira, D.A. de M.	126	Wang, Z.	195, 368
Phongpaew, T.	205	Wei, H.	221
Ponsree, K.	205	Wu, F.	688
Pradana, Z.H.	527	Wu, Q.	256
Qi, H.	695	Xiao, Y.S.	632
Qiu, X.	427	Xie, H.	688

Xin, J.	458	Zeng, X.	305
Xu, G.	282	Zhang, J.	351
Xu, H.	590	Zhang, X.	605
Xu, J.	406	Zhang, X.-d.	622
Xu, L.	416	Zhang, Yulin	703
Xu, X.	458	Zhang, Yu	709
Xu, Z.	481	Zhang, Z.	695
Yamshchikov, I.P.	502, 511	Zhao, D.	195
Yan, J.	688	Zhao, F.	368
Yan, N.	385	Zhao, H.	695
Yan, W.	228	Zhao, Y.	296, 393, 442
Yang, J.	195	Zhong, Q.	632
Yang, W.	228	Zhou, B.	256
Yang, Y.	676	Zhou, H.-j.	622
Yin, M.	49	Zhou, N.	66, 77, 317
Yin, W.	695	Zhou, W.	221
Yin, Z.	416	Zhou, X.	351
Yu, F.	16	Zhou, Y.	40, 416, 427
Yu, X.	296	Zhu, H.	688
Zeng, M.	614		

This page intentionally left blank