Available online at www.sciencedirect.com
SCIENCE DIRECT®

Computers and Chemical Engineering 27 (2003) 293–311

Computers & Chemical Engineering

www.elsevier.com/locate/compchemeng

A review of process fault detection and diagnosis Part I: Quantitative model-based methods

Venkat Venkatasubramanian [a,*], Raghunathan Rengaswamy [b,*], Kewen Yin [c], Surya N. Kavuri [d]

[a] *Laboratory for Intelligent Process Systems, School of Chemical Engineering, Purdue University, West Lafayette, IN 47907, USA*
[b] *Department of Chemical Engineering, Clarkson University, Potsdam, NY 13699-5705, USA*
[c] *Department of Wood and Paper Science, University of Minnesota, St. Paul, MN 55108, USA*
[d] *BP, Houston, TX, USA*

Received 12 February 2001; accepted 22 April 2002

Abstract

Fault detection and diagnosis is an important problem in process engineering. It is the central component of abnormal event management (AEM) which has attracted a lot of attention recently. AEM deals with the timely detection, diagnosis and correction of abnormal conditions of faults in a process. Early detection and diagnosis of process faults while the plant is still operating in a controllable region can help avoid abnormal event progression and reduce productivity loss. Since the petrochemical industries lose an estimated 20 billion dollars every year, they have rated AEM as their number one problem that needs to be solved. Hence, there is considerable interest in this field now from industrial practitioners as well as academic researchers, as opposed to a decade or so ago. There is an abundance of literature on process fault diagnosis ranging from analytical methods to artificial intelligence and statistical approaches. From a modelling perspective, there are methods that require accurate process models, semi-quantitative models, or qualitative models. At the other end of the spectrum, there are methods that do not assume any form of model information and rely only on historic process data. In addition, given the process knowledge, there are different search techniques that can be applied to perform diagnosis. Such a collection of bewildering array of methodologies and alternatives often poses a difficult challenge to any aspirant who is not a specialist in these techniques. Some of these ideas seem so far apart from one another that a non-expert researcher or practitioner is often left wondering about the suitability of a method for his or her diagnostic situation. While there have been some excellent reviews in this field in the past, they often focused on a particular branch, such as analytical models, of this broad discipline. The basic aim of this three part series of papers is to provide a systematic and comparative study of various diagnostic methods from different perspectives. We broadly classify fault diagnosis methods into three general categories and review them in three parts. They are quantitative model-based methods, qualitative model-based methods, and process history based methods. In the first part of the series, the problem of fault diagnosis is introduced and approaches based on quantitative models are reviewed. In the remaining two parts, methods based on qualitative models and process history data are reviewed. Furthermore, these disparate methods will be compared and evaluated based on a common set of criteria introduced in the first part of the series. We conclude the series with a discussion on the relationship of fault diagnosis to other process operations and on emerging trends such as hybrid blackboard-based frameworks for fault diagnosis.

Keywords: Fault detection; Diagnosis; Process safety

1. Introduction

The discipline of process control has made tremendous advances in the last three decades with the advent of computer control of complex processes. Low-level control actions such as opening and closing valves, called regulatory control, which used to be performed by

* Corresponding authors. Tel.: +1-765-494-0734; fax: +1-765-494-0805 (V. Venkatsubramanian), Tel.: +1-315-268-4423; fax: +1-315-268-6654 (R. Rengaswamy).
E-mail addresses: venkat@ccn.purdue.edu (V. Venkatasubramanian), raghu@clarkson.edu (R. Rengaswamy).

0098-1354/03/$ - see front matter © 2002 Published by Elsevier Science Ltd.
PII: S0098-1354(02)00160-6

human operators are now routinely performed in an automated manner with the aid of computers with considerable success. With progress in distributed control and model predictive control systems, the benefits to various industrial segments such as chemical, petrochemical, cement, steel, power and desalination industries have been enormous. However, a very important control task in managing process plants still remains largely a manual activity, performed by human operators. This is the task of responding to abnormal events in a process. This involves the timely detection of an abnormal event, diagnosing its causal origins and then taking appropriate supervisory control decisions and actions to bring the process back to a normal, safe, operating state. This entire activity has come to be called Abnormal Event Management (AEM), a key component of supervisory control.

However, this complete reliance on human operators to cope with such abnormal events and emergencies has become increasingly difficult due to several factors. It is difficult due to the broad scope of the diagnostic activity that encompasses a variety of malfunctions such as process unit failures, process unit degradation, parameter drifts and so on. It is further complicated by the size and complexity of modern process plants. For example, in a large process plant there may be as many as 1500 process variables observed every few seconds (Bailey, 1984) leading to information overload. In addition, often the emphasis is on quick diagnosis which poses certain constraints and demands on the diagnostic activity. Furthermore, the task of fault diagnosis is made difficult by the fact that the process measurements may often be insufficient, incomplete and/or unreliable due to a variety of causes such as sensor biases or failures.

Given such difficult conditions, it should come as no surprise that human operators tend to make erroneous decisions and take actions which make matters even worse, as reported in the literature. Industrial statistics show that about 70% of the industrial accidents are caused by human errors. These abnormal events have significant economic, safety and environmental impact. Despite advances in computer-based control of chemical plants, the fact that two of the worst ever chemical plant accidents, namely Union Carbide's Bhopal, India, accident and Occidental Petroleum's Piper Alpha accident (Lees, 1996), happened in recent times is a troubling development. Another major recent incident is the explosion at the Kuwait Petrochemical's Mina Al-Ahmedi refinery in June of 2000, which resulted in about 100 million dollars in damages.

Further, industrial statistics have shown that even though major catastrophes and disasters from chemical plant failures may be infrequent, minor accidents are very common, occurring on a day to day basis, resulting in many occupational injuries, illnesses, and costing the society billions of dollars every year (Bureau of Labor Statistics, 1998; McGraw-Hill Economics, 1985; National Safety Council, 1999). It is estimated that the petrochemical industry alone in the US incurs approximately 20 billion dollars in annual losses due to poor AEM (Nimmo, 1995). The cost is much more when one includes similar situations in other industries such as pharmaceutical, specialty chemicals, power and so on. Similar accidents cost the British economy up to 27 billion dollars every year (Laser, 2000).

Thus, here is the next grand challenge for control engineers. In the past, the control community showed how regulatory control could be automated using computers and thereby removing it from the hands of human operators. This has led to great progress in product quality and consistency, process safety and process efficiency. The current challenge is the automation of AEM using intelligent control systems, thereby providing human operators the assistance in this most pressing area of need. People in the process industries view this as the next major milestone in control systems research and application.

The automation of process fault detection and diagnosis forms the first step in AEM. Due to the broad scope of the process fault diagnosis problem and the difficulties in its real time solution, various computer-aided approaches have been developed over the years. They cover a wide variety of techniques such as the early attempts using fault trees and digraphs, analytical approaches, and knowledge-based systems and neural networks in more recent studies. From a modelling perspective, there are methods that require accurate process models, semi-quantitative models, or qualitative model. At the other end of the spectrum, there are methods that do not assume any form of model information and rely only on process history information. In addition, given the process knowledge, there are different search techniques that can be applied to perform diagnosis. Such a collection of bewildering array of methodologies and alternatives often pose a difficult challenge to any aspirant who is not a specialist in these techniques. Some of these ideas seem so far apart from one another that a non-expert researcher or practitioner is often left wondering about the suitability of a method for his or her diagnostic situation. While there have been some excellent reviews in this filed in the past, they often focused on a particular branch, such as analytical models, of this broad discipline.

The basic aim of this three part series of papers is to provide a systematic and comparative study of various diagnostic methods from different perspectives. We broadly classify fault diagnosis methods into three general categories and review them in three parts. They are quantitative model based methods, qualitative model based methods, and process history based methods. We review these different approaches and attempt to present a perspective showing how these different methods relate to and differ from each other. While discussing these various methods we will also try to point out important assumptions, drawbacks as well as advantages that are not stated explicitly and are difficult to gather. Due to the broad nature of this exercise it is not possible to discuss every method in all its detail. Hence the intent is to provide the reader with the general concepts and lead him or her on to literature that will be a good entry point into this field.

In the first part of the series, the problem of fault diagnosis is introduced and fault diagnosis approaches based on quantitative models are reviewed. In the following two parts, fault diagnostic methods based on qualitative models and process history data are reviewed. Further, these disparate methods will be compared and evaluated based on a common set of desirable characteristics for fault diagnostic classifiers introduced in this paper. The relation of fault diagnosis to other process operations and a discussion on future directions are presented in Part III.

By way of introduction, we first address the definitions and nomenclature used in the area of process fault diagnosis. The term *fault* is generally defined as a departure from an acceptable range of an observed variable or a calculated parameter associated with a process (Himmelblau, 1978). This defines a fault as a process abnormality or symptom, such as high temperature in a reactor or low product quality and so on. The underling cause(s) of this abnormality, such as a failed coolant pump or a controller, is(are) called the *basic event(s)* or the *root cause(s)*. The basic event is also referred to as a *malfunction* or a *failure*. Since one can view the task of diagnosis as a classification problem, the diagnostic system is also referred to as a diagnostic classifier. Fig. 1 depicts the components of a general fault diagnosis framework. The figure shows a controlled process system and indicates the different sources of failures in it. In general, one has to deal with three classes of failures or malfunctions as described below:

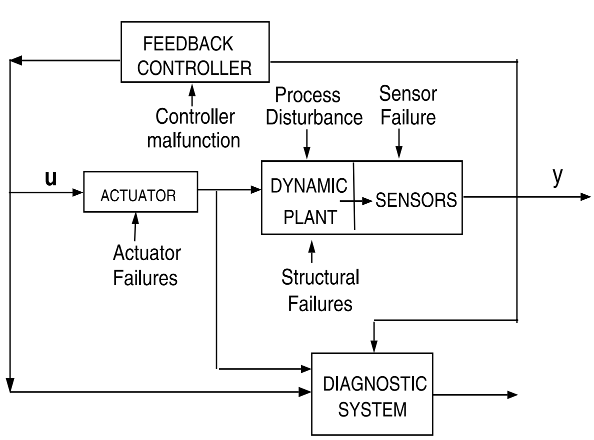

Fig. 1. A general diagnostic framework.

1.1. Gross parameter changes in a model

In any modelling, there are processes occurring below the selected level of detail of the model. These processes which are not modelled are typically lumped as parameters and these include interactions across the system boundary. Parameter failures arise when there is a disturbance entering the process from the environment through one or more exogenous (independent) variables. An example of such a malfunction is a change in the concentration of the reactant from its normal or steady state value in a reactor feed. Here, the concentration is an exogenous variable, a variable whose dynamics is not provided with that of the process. Another example is the change in the heat transfer coefficient due to fouling in a heat exchanger.

1.2. Structural changes

Structural changes refer to changes in the process itself. They occur due to hard failures in equipment. Structural malfunctions result in a change in the information flow between various variables. To handle such a failure in a diagnostic system would require the removal of the appropriate model equations and restructuring the other equations in order to describe the current situation of the process. An example of a structural failure would be failure of a controller. Other examples include a stuck valve, a broken or leaking pipe and so on.

1.3. Malfunctioning sensors and actuators

Gross errors usually occur with actuators and sensors. These could be due to a fixed failure, a constant bias (positive or negative) or an out-of range failure. Some of the instruments provide feedback signals which are essential for the control of the plant. A failure in one of the instruments could cause the plant state variables to deviate beyond acceptable limits unless the failure is detected promptly and corrective actions are accomplished in time. It is the purpose of diagnosis to quickly detect any instrument fault which could seriously degrade the performance of the control system.

Outside the scope of fault diagnosis are unstructured uncertainties, process noise and measurement noise. Unstructured uncertainties are mainly faults that are not modelled a priori. Process noise refers to the mismatch between the actual process and the predictions from model equations, whereas, measurement noise refers to high frequency additive component in the sensor measurements.

In this series of review papers, we will provide a review of the various techniques that have been proposed to solve the problem of fault detection and diagnosis. We classify the techniques as quantitative model based, qualitative model based and process history based approaches. Under the quantitative model based approaches, we will review techniques that use analytical redundancy to generate residuals that can be used for isolating process failures. We will discuss residual generation through diagnostic observers, parity relations, Kalman filters and so on. Under the qualitative model based approaches, we review the signed directed graph (SDG), Fault Trees, Qualitative Simulation (QSIM), and Qualitative Process Theory (QPT) approaches to fault diagnosis. Further, we also classify diagnostic search strategies as being topographic or symptomatic searches. Under process history based approaches we will discuss both qualitative approaches such as expert systems and qualitative trend analysis (QTA) techniques and quantitative approaches such as neural networks, PCA and statistical classifiers.

We believe that there have been very few articles that comprehensively review the field of fault diagnosis considering all the different types of techniques that have been discussed in this series of review papers. Most of the review papers such as the one by Frank, Ding, and Marcu (2000) seem to focus predominantly on model based approaches. For example, in the review by Frank et al., a detailed description of various types of analytical model based approaches is presented. The robustness issues in fault detection, optimized generation of residuals and generation of residuals for nonlinear systems are some of the issues that have been addressed in a comprehensive manner. There are a number of other review articles that fall under the same category. A brief review article that is more representative of all the available fault diagnostic techniques has been presented by Kramer and Mah (1993). This review deals with data validation, rectification and fault diagnosis issues. The fault diagnosis problem is viewed as consisting of feature extraction and classification stages. This view of fault diagnosis has been generalized in our review as the transformations that measurements go through before a final diagnostic decision is attained. The classification stage is examined by Kramer and Mah as falling under three main categories. (i) pattern recognition, (ii) model-based reasoning and (iii) model-matching. Under pattern recognition, most of the process history based methods are discussed; under model-based reasoning most of the qualitative model based techniques are discussed; and symptomatic search techniques using different model forms are discussed under model matching techniques.

Closely associated with the area of fault detection and diagnosis is the research area of gross error detection in sensor data and the subsequent validation. Gross error detection or sensor validation refers to the identification of faulty or failed sensors in the process. Data reconciliation or rectification is the task of providing estimates for the true values of sensor readings. There has been considerable work done in this area and there have also been review papers and books written on this area. Hence, we do not provide a review of this field in this series of papers. However, as mentioned before, fault diagnosis includes sensor failures also in its scope and hence data validation and rectification is a specific case of a more general fault diagnosis problem (Kramer & Mah, 1993).

The rest of this first part of the review is organized as follows. In the next section, we propose a list of ten desirable characteristics that one would like a diagnostic system to possess. This list would help us assess the various approaches against a common set of criteria. In Section 3, we discuss the transformations of data that take place during the process of diagnostic decision-making. This discussion lays down the framework for analyzing the various diagnostic approaches in terms of their knowledge and search components. In Section 4, a classification of fault diagnosis methods is provided. In Section 5, diagnosis methods based on quantitative models are discussed in detail.

2. Desirable characteristics of a fault diagnostic system

In the last section, the general problem of fault diagnosis was presented. In order to compare various diagnostic approaches, it is useful to identify a set of desirable characteristics that a diagnostic system should possess. Then the different approaches may be evaluated against such a common set of requirements or standards. Though these characteristics will not usually be met by any single diagnostic method, they are useful to benchmark various methods in terms of the a priori information that needs to be provided, reliability of solution, generality and efficiency in computation etc. In this context, one needs to understand the important concepts, completeness and resolution, before proceeding to the characteristics of a good diagnostic classifier. Whenever an abnormality occurs in a process, a general diagnostic classifier would come up with a set of hypotheses or faults that explains the abnormality. Completeness of a diagnostic classifier would require the actual fault(s) to be a subset of the proposed fault set. Resolution of a diagnostic classifier would require the fault set to be as minimal as possible. Thus, there is a trade-off between completeness and resolution. The trade-off is in the accuracy of predictions. These two concepts would recur whenever different classifier designs are compared. The following presents a set of desirable characteristics one would like the diagnostic system to possess.

2.1. Quick detection and diagnosis

The diagnostic system should respond quickly in detecting and diagnosing process malfunctions. However, quick response to failure diagnosis and tolerable performance during normal operation are two conflicting goals (Willsky, 1976). A system that is designed to detect a failure (particularly abrupt changes) quickly will be sensitive to high frequency influences. This makes the system sensitive to noise and can lead to frequent false alarms during normal operation, which can be disruptive. This is analogous to the trade-off between robustness and performance noticed in the control literature.

2.2. Isolability

Isolability is the ability of the diagnostic system to distinguish between different failures. Under ideal conditions free of noise and modelling uncertainties, this amounts to saying that the diagnostic classifier should be able to generate output that is orthogonal to faults that have not occurred. Of course the ability to design isolable classifiers depends to a great extent on the process characteristics. There is also a trade-off between isolability and the rejection of modelling uncertainties. Most of the classifiers work with various forms of redundant information and hence there is only a limited degree of freedom for classifier design. Due to this, a classifier with high degree of isolability would usually do a poor job in rejecting modelling uncertainties and vice versa.

2.3. Robustness

One would like the diagnostic system to be robust to various noise and uncertainties. One would like its performance to degrade gracefully instead of failing totally and abruptly. Robustness precludes deterministic isolability tests where the thresholds are placed close to zero. In the presence of noise, these thresholds may have to be chosen conservatively. Thus, as noted earlier, robustness needs are to be balanced with those of performance.

2.4. Novelty identifiability

One of the minimal requirements of a diagnostic system is to be able to decide, given current process conditions, whether the process is functioning normally or abnormally, and if abnormal, whether the cause is a known malfunction or an unknown, novel, malfunction. This criterion is known as novelty identifiability. In general, sufficient data may be available to model the normal behavior of the process. However, one typically does not have such historic process data available for modelling the abnormal regions satisfactorily (of course, if one has access to a good dynamic model of the process, then generating such data is much easier). Only a few data patterns may be available covering portions of the abnormal region. Thus, it is possible that much of the abnormal operations region may not have been modelled adequately. This will pose serious challenges in achieving novelty identifiability. Even under these difficult conditions, one would like the diagnostic system to be able to recognize the occurrence of novel faults and not misclassify them as one of the other known malfunctions or as normal operation.

2.5. Classification error estimate

An important practical requirement for a diagnostic system is in building the user's confidence on its reliability. This could be greatly facilitated if the diagnostic system could provide a priori estimate on classification error that can occur. Such error measures would be useful to project confidence levels on the diagnostic decisions by the system giving the user a better feel for the reliability of the recommendations by the system.

2.6. Adaptability

Processes in general change and evolve due to changes in external inputs or structural changes due to retrofitting and so on. Process operating conditions can change not only due to disturbances but also due to changing environmental conditions such as changes in production quantities with changing demands, changes in the quality of raw material etc. Thus the diagnostic system should be adaptable to changes. It should be possible to gradually develop the scope of the system as new cases and problems emerge, as more information becomes available.

2.7. Explanation facility

Besides the ability to identify the source of malfunction, a diagnostic system should also provide explanations on how the fault originated and propagated to the current situation. This is a very important factor in designing on-line decision support systems. This requires the ability to reason about cause and effect relationships in a process. A diagnostic system has to justify its recommendations so that the operator can accordingly evaluate and act using his/her experience. One would like the diagnostic system to not only justify why certain hypotheses were proposed but also explain why certain other hypotheses were not proposed.

2.8. Modelling requirements

The amount of modelling required for the development of a diagnostic classifier is an important issue. For fast and easy deployment of real-time diagnostic classifiers, the modelling effort should be as minimal as possible.

2.9. Storage and computational requirements

Usually, quick real-time solutions would require algorithms and implementations which are computationally less complex, but might entail high storage requirements. One would prefer a diagnostic system that is able to achieve a reasonable balance on these two competing requirements.

2.10. Multiple fault identifiability

The ability to identify multiple faults is an important but a difficult requirement. It is a difficult problem due to the interacting nature of most faults. In a general nonlinear system, the interactions would usually be synergistic and hence a diagnostic system may not be able to use the individual fault patterns to model the combined effect of the faults. On the other hand, enumerating and designing separately for various multiple fault combinations would become combinatorially prohibitive for large processes.

3. Transformations of measurements in a diagnostic system

To attempt a comparative study of various diagnostic methods it is helpful to view them from different perspectives. In this sense, it is important to identify the various transformations that process measurements go through before the final diagnostic decision is made. Two important components in the transformations are the a priori process knowledge and the search technique used. Hence, one can discuss diagnostic methods from these two perspectives. Also, one can view diagnostic methods based on different solution philosophies like knowledge-based systems, pattern recognition, and analytical model-based methods. These methods have distinct diagnostic architectures and utilize different combinations of a priori knowledge and search techniques. Though these are fundamentally different viewpoints, it is hard to draw clear demarcation between these viewpoints and hence a certain amount of overlap is unavoidable.

In general, one can view the diagnostic decision-making process as a series of transformations or mappings on process measurements. Fig. 2 shows the various transformations that process data go through during diagnosis. The *measurement* space is a space of measurements x_1, x_2,...,x_N with no a priori problem knowledge relating these measurements. These are the input to the diagnostic system. The *feature space* is a space of points $y = (y_1,...,y_i)$ where y_i is the i^{th} feature obtained as a function of the measurements by utilizing a priori problem knowledge. Here, the measurements are analyzed and combined with the aid of a priori process knowledge to extract useful features about the process behavior to aid diagnosis. The mapping from the *feature space* to *decision space* is usually designated to meet some objective function (such as minimizing the misclassification). This transformation is achieved by either using a discriminant function or in some cases using simple threshold functions. The decision space is a space of points $d = [d_1,...,d_K]$, where K is the number of decision variables, obtained by suitable transformations of the feature space. The *class space* is a set of integers $c = [c_1, \ldots, c_M]$, where M is the number of failure classes, indexing the failure classes indicating categorically to which failure class (or classes) including normal region a given measurement pattern belongs. The class space is thus the final interpretation of the diagnostic system delivered to the user. The transformations from decision space to class space is again performed using either threshold functions, template matching or symbolic reasoning as the case may be.

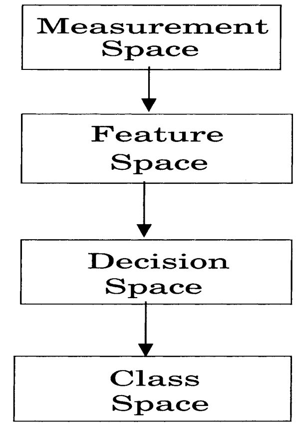

Fig. 2. Transformations in a diagnostic system.

There are two ways of developing the feature space from the measurement space, namely, feature selection and feature extraction. In feature selection, one simply selects a few important measurements of the original measurement space. Feature extraction is a procedure that facilitates, through the use of prior knowledge of the problem, a transformation of the measurement space into a space of fewer dimensions. For example, if a relationship is known to exist between the samples of one measurement and the samples of another measurement, feature extraction is concerned with identifying this relationship. Once this relationship is identified, rather than having two sets of parameters characterizing the two dimensions, it may be possible to represent them with a single set of parameters.

To explain these transformations more clearly, let us consider a simple example. Let there be four sensors x_1, x_2, x_3, x_4 and let there be two fault classes c_1 and c_2 that need to be distinguished. Let us suppose further that fault 1 affects sensors 1 and 2 and fault 2 affects sensors 2 and 3. Let us also suppose x_{1ss}, x_{2ss}, x_{3ss}, x_{4ss} are the steady-state values of these sensors. In this case, one simple transformation to form a feature space would be to drop sensor measurement x_4. Hence the feature space would be $[y_1 \quad y_2 \quad y_3]=[x_1 \quad x_2 \quad x_3]$. Now, there are different ways in which one can transform this feature space to a decision space. One transformation could be a simple threshold function to form the decision space $[d_1 \quad d_2 \quad d_3]$. The threshold function would be: If ABS $(y_i - x_{iss}) > T$ then $d_i = 1$ else $d_i = 0$. The final transformation would be from the decision space to class space $[c_1 \quad c_2]$ and it can be performed by symbolic logic. For example, "IF (d_1 AND d_2) Then c_1 and IF (d_2 AND d_3) Then c_2" would be a valid logic for this transformation.

To provide another example, consider the Bayes classifier. The Bayes classifier for a two class problem, assuming Gaussian density function for the classes, is developed as follows (Fukunaga, 1972): measurements x are first transformed using a priori model information into features y. These features are then transformed into the decision space—which is a set of real numbers indexed by fault classes. The real number corresponding to fault class i is the distance d_i of feature y from the mean m_i of class i scaled by the covariance Σ_i of class i. For a two class problem, we have:

$$d_1 = (y - m_1)^T \sum\nolimits_1^{-1} (y - m_1)$$

$$d_2 = (y - m_2)^T \sum\nolimits_2^{-1} (y - m_2)$$

where $[d_1, d_2]$ spans the decision space as x spans the measurement space. A discriminant function h maps the decision space to class space (with a priori probabilities for the classes being the same).

$$h = d_1 - d_2$$

$h < \delta, x$ belongs to class I

$h > \delta, x$ belongs to class II

$\delta = \log\left(\dfrac{\left|\Sigma_2\right|}{\left|\Sigma_1\right|}\right)$ where δ is the threshold of the classifier.

In a neural network based classifier, input nodes represent the measurement space. Hidden nodes correspond to the feature space. Output nodes map the feature space to the decision space. An interpretation of the outputs gives the mapping to class space. In analytical model approaches (Gertler, 1991; Frank, 1990), the residuals from different models define the feature space. From these, one may compute the likelihood of occurrence of various fault classes, which in turn can be used to specify which fault classes are actually present.

In most cases, the decision space and class space have the same dimension. Still, it would be preferable to maintain separate decision and class spaces because in some cases one might not be able to force the diagnostic classifier to come up with crisp solutions. Consider neural networks as a diagnostic classifier as an example. The output nodes represent the decision space. One would still need a crisp solution based on some kind of threshold function for the final interpretation to the user.

The basic assumption in transforming the measurement space to feature space is that the features cluster better in the feature space than the measurements do in the measurement space, thus facilitating improved classification or better discrimination. The advantage one gains in developing a feature space is the reduction in the complexity of the discriminant function. The transformation from measurement to feature space is done using a priori process knowledge, whereas, the transformation from feature space to decision space is implemented as a search or learning algorithm. The decision space is usually mapped to the class space using simple threshold functions. If the process plant under consideration is a well understood one, then one has powerful a priori knowledge to work with. Such a strong understanding would help one design an effective mapping from the measurement space to a feature space that has discriminating features. By simplifying the problem using a priori knowledge to form a powerful feature space, the burden on the search/learning algorithm can be reduced greatly.

Thus, a priori process knowledge plays a crucial role in diagnostic decision-making through the mappings. There are different kinds of a priori knowledge that may be available for this purpose depending on the user's understanding of the process. For example, a priori knowledge could be available in the form of invariant relationships between sensor outputs and actuator inputs such as material and energy balances in a process system (Mehra & Peschon, 1971; Kramer, 1987) or it could be a suitably transformed model of the process, resulting in a bank of filters (Willsky, 1976; Gertler, 1991; Frank, 1990). A priori knowledge can also be in the form of distribution information in the measurement space (Hoskins & Himmelblau, 1988).

By pursuing this theme of transformations between spaces in diagnostic reasoning, a comparative study on various approaches to fault diagnosis may be made through an analysis of the different forms of a priori knowledge used to develop the feature space and the search strategies used to arrive at the mapping to decision/class space. This would shed some light on how the different approaches relate to and differ from one another within this framework of spatial mappings.

4. Classification of diagnostic algorithms

As discussed earlier two of the main components in a diagnosis classifier are: (i) the type of knowledge and (ii) the type of diagnostic search strategy. Diagnostic search strategy is usually a very strong function of the knowledge representation scheme which in turn is largely influenced by the kind of a priori knowledge available. Hence, the type of a priori knowledge used is the most important distinguishing feature in diagnostic systems. In this three part review paper we classify the diagnostic systems based on the a priori knowledge used.

The basic a priori knowledge that is needed for fault diagnosis is the set of failures and the relationship between the observations (symptoms) and the failures. A diagnostic system may have them explicitly (as in a table lookup), or it may be inferred from some source of domain knowledge. The a priori domain knowledge may be developed from a fundamental understanding of the process using first-principles knowledge. Such knowledge is referred to as deep, causal or model-based knowledge (Milne, 1987). On the other hand, it may be gleaned from past experience with the process. This knowledge is referred to as shallow, compiled, evidential or process history-based knowledge.

The model-based a priori knowledge can be broadly classified as qualitative or quantitative. The model is usually developed based on some fundamental understanding of the physics of the process. In quantitative models this understanding is expressed in terms of mathematical functional relationships between the inputs and outputs of the system. In contrast, in qualitative model equations these relationships are expressed in terms of qualitative functions centered around different units in a process.

In contrast to the model-based approaches where a priori knowledge about the model (either quantitative or qualitative) of the process is assumed, in process history based methods only the availability of large amount of historical process data is assumed. There are different ways in which this data can be transformed and presented as a priori knowledge to a diagnostic system. This is known as the feature extraction process from the process history data, and is done to facilitate later diagnosis. This extraction process can mainly proceed as either quantitative or qualitative feature extraction. In quantitative feature extraction one can perform either a statistical or non-statistical feature extraction. This classification of diagnostic systems is shown in Fig. 3. In this paper, we focus on diagnostic systems that are built on quantitative models. The remaining two parts will focus on diagnostic methods based on qualitative models and process history data.

A bit of clarification on the classification is in order at this point. It is clear that all models need data for estimating some of the parameters in the model and all the methods based on process data need to extract some form of a model to perform fault diagnosis. The classification of quantitative, qualitative and process history in our view provides a classification in terms of the manner in which these methods approach the problem of fault diagnosis. As an example, though the models for observers (classified under quantitative approach) are based on input–output data, the use of these models in generating diagnostic results largely follows a quantitative approach. However, a qualitative approach such as QSIM, again based on a model (in fact, a first principles quantitative model), uses a distinctly qualitative framework for diagnostic explanation generation. Similarly, neural networks approaches for fault diagnosis have largely been approached from a pattern recognition point of view and hence we have classified these approaches under process history based methods, though they are directly related to state–space models. We believe that in spite of the overlap this is a good classification for fault detection and isolation (FDI) strategies.

5. Quantitative model-based approaches

This section reviews quantitative model-based fault diagnosis methods. The concept of analytical redundancy is introduced first, followed by a description of discrete dynamic system with linear models. The most frequently used FDI approaches, including diagnostic observers, parity relations, Kalman filters and parameter estimation are outlined. The recent effort of generating enhanced residuals to facilitate the fault isolation procedure is discussed. We will discuss the principles behind these methods, summarize their main applications, comment on their advantages, deficiencies and tradeoffs, and review their most recent advances in process monitoring and fault diagnosis. Our purpose is to provide an overview of the basic concepts in model-based fault detection. For the sake of brevity, we have included only some of the many techniques and therefore the references listed in this review paper are by no means exhaustive. However, we believe that they are good resources to the interested readers for further study.

Further, most of the work on quantitative model-based approaches have been based on general input–output and state–space models as discussed below. However, there are a wide variety of quantitative model types that have been considered in fault diagnosis such

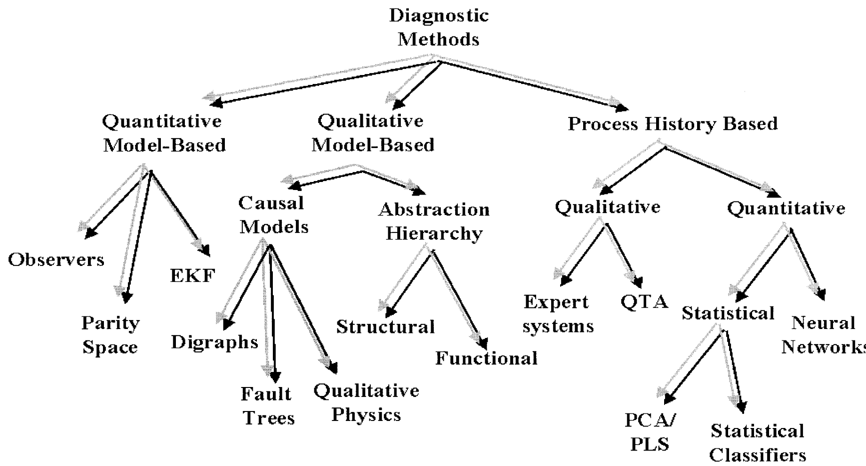

Fig. 3. Classification of diagnostic algorithms.

as first-principles models, frequency response models and so on. The first-principles models (also classified as macroscopic transport phenomena model (Himmelblau, 1978)) have not been very popular in fault diagnosis studies because of the computational complexity in utilizing these models in real-time fault diagnostic systems and the difficulty in developing these models. The most important class of models that have been heavily investigated in fault diagnosis studies are the input–output or state–space models and hence the focus is on these types of models.

5.1. Analytical redundancy

In the area of automatic control, change/fault detection problems are known as model-based FDI. Relying on an explicit model of the monitored plant, all model-based FDI methods (and many of the statistical diagnosis methods) require two steps. The first step generates inconsistencies between the actual and expected behavior. Such inconsistencies, also called *residuals*, are 'artificial signals' reflecting the potential faults of the system. The second step chooses a decision rule for diagnosis.

The check for inconsistency needs some form of redundancy. There are two types of redundancies, hardware redundancy and analytical redundancy. The former requires redundant sensors. It has been utilized in the control of such safety-critical systems as aircraft space vehicles and nuclear power plants. However, its applicability is limited due to the extra cost and additional space required. On the other hand, analytical redundancy (also termed functional, inherent or artificial redundancy) is achieved from the functional dependence among the process variables and is usually provided by a set of algebraic or temporal relationships among the states, inputs and the outputs of the system. According to how the redundancy is accomplished, analytical redundancy can be further classified into two categories (Basseville, 1988; Chow & Willsky, 1984; Frank, 1990), direct and temporal.

A direct redundancy is accomplished from algebraic relationships among different sensor measurements. Such relationship are useful in computing the value of a sensor measurement from measurements of other sensors. The computed value is then compared with the measured value from that sensor. A discrepancy indicates that a sensor fault may have occurred.

A temporal redundancy is obtained from differential or difference relationships among different sensor outputs and actuator inputs. With process input and output data, temporal redundancy is useful for sensor and actuator fault detection.

A general scheme of using analytical redundancy in diagnostic systems is given in Fig. 4. The essence of analytical redundancy in fault diagnosis is to check the actual system behavior against the system model for consistency. Any inconsistency expressed as residuals, can be used for detection and isolation purposes. The residuals should be close to zero when no fault occurs but show 'significant' values when the underlying system changes. The generation of the diagnostic residuals requires an explicit mathematical model of the system. Either a model derived analytically using first principles or a black-box model obtained empirically may be used.

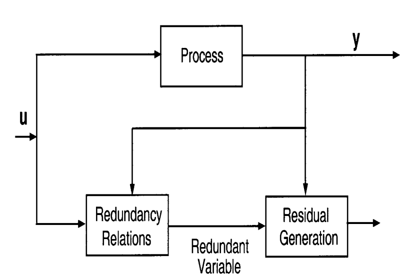

Fig. 4. General scheme for using analytical redundancy.

Also, statistical methods are often required for the decision making.

The first principles models are obtained based on a physical understanding of the process. In a chemical engineering process, mass, energy and momentum balances as well as constitutive relationships (such as equations of state) are used in the development of model equations. In the past, models developed from first principles were seldom used in process control and fault diagnosis mainly because of their complexity. In addition, the chemical engineering processes are often nonlinear, which makes the design of fault diagnosis procedures more difficult. However, owing to the availability of better and faster computers and the improved understanding of nonlinear controller design and synthesis, this situation is improving.

The problem of fault diagnosis is one of identifying the state of a process based on its behavior. The behavior of a process is monitored through its sensor outputs and actuator inputs. When faults occur, they change the relationship among these observed variables and therefore result in nonzero residuals. Most of the FDI methods use discrete black-box plant models such as input–output or state–space models and assume linearity of the plant. The main difference between the first principles and the black-box models is that the parameters in the former bear certain physical meanings, which can be very useful in the diagnostic procedure or the controller design.

Although dynamic systems are continuous processes, all the diagnostic tools use sampled data, and hence only discrete models are included herein. However the basic concepts, if not the detailed analysis, carry over to continuous models. In addition, most of the model-based approaches assume system linearity. Their application to a non-linear system requires a model linearization around the operating point.

Consider a system with m inputs and k outputs. Let $\mathbf{u}(t)=[u_1(t)\ldots u_m(t)]^T$ be the process inputs and $\mathbf{y}(t)=[y_1(t)\ \ldots\ y_k(t)]^T$ be the process outputs, where t denotes the discrete time. The basic system model in the state–space form is

$$\mathbf{x}(t+1)=\mathbf{A}\mathbf{x}(t)+\mathbf{B}\mathbf{u}(t)$$

$$\mathbf{y}(t)=\mathbf{C}\mathbf{x}(t)+\mathbf{D}\mathbf{u}(t) \tag{1}$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ are parameter matrices with appropriate dimensions; $\mathbf{x}(t)$ denotes the n dimensional state vector.

The same system can be expressed in the input–output form

$$\mathbf{H}(z)\mathbf{y}(t)=\mathbf{G}(z)\mathbf{u}(t) \tag{2}$$

where $\mathbf{H}(z)$ and $\mathbf{G}(z)$ are polynomial matrices in z^{-1}, (the backward-shift operator), $\mathbf{H}(z)$ is diagonal; $\mathbf{H}(z)$ and $\mathbf{G}(z)$ are of the form

$$\mathbf{H}(z)=\mathbf{I}+\mathbf{H}_1z^{-1}+\mathbf{H}_2z^{-2}+\ldots+\mathbf{H}_nz^{-n}$$

$$\mathbf{G}(z)=\mathbf{G}_0+\mathbf{G}_1z^{-1}+\mathbf{G}_2z^{-2}+\ldots+\mathbf{G}_nz^{-n}$$

Process models (1) and (2) describe an ideal situation where there are no faults or any form of disturbances and/or noise. Faults in the state–space framework are usually modelled as (Gertler, 1991, 1993)

$$\mathbf{x}(t+1)=\mathbf{A}\mathbf{x}(t)+\mathbf{B}\mathbf{u}(t)+\mathbf{E}\mathbf{p}(t)$$

$$\mathbf{y}(t)=\mathbf{C}\mathbf{x}(t)+\mathbf{D}\mathbf{u}(t)+\mathbf{E}'\mathbf{p}(t)+\mathbf{q}(t) \tag{3}$$

where input commands $\mathbf{u}(t)$ and measured outputs $\mathbf{y}(t)$ are both observable. Included in $\mathbf{p}(t)$ are actuator faults, certain plant faults, disturbances as well as input sensor faults. $\mathbf{q}(t)$ represents output sensor faults. In the input–output framework, Eq. (2) is replaced with

$$\mathbf{H}(z)\mathbf{y}(t)=\mathbf{G}(z)\mathbf{u}(t)+\mathbf{H}(z)\mathbf{q}(t)+\mathbf{F}(z)\mathbf{p}(t) \tag{4}$$

where $\mathbf{q}(t)$ and $\mathbf{p}(t)$ are as defined above.

When developing models for the real situation, we need to distinguish between two different forms of faults, additive and multiplicative, both of which result in additional terms in the process model (2). Multiplicative faults lead to changes in the parameters (i.e. in matrices $\mathbf{H}(z)$ and $\mathbf{G}(z)$) and depend on the actual values of the observed variables. With multiplicative faults, the model (2) changes to (Gertler, 1992)

$$(\mathbf{H}(z)+\Delta\mathbf{H}(z))\mathbf{y}(t)=(\mathbf{G}(z)+\Delta\mathbf{G}(z))\mathbf{u}(t) \tag{5}$$

Additive faults, on the other hand, appear as additional terms in the process model (2) and are independent of the values of the observed variables. Unknown disturbances are included in Eq. (2) as

$$\mathbf{H}(z)\mathbf{y}(t)=\mathbf{G}(z)\mathbf{u}(t)+\mathbf{H}(z)\mathbf{q}(t)+\mathbf{F}(z)\mathbf{p}(t)+\mathbf{K}(z)\omega(t) \tag{6}$$

where $\mathbf{q}(t)$ are the sensor faults; $\mathbf{p}(t)$ represent the actuator faults; and $\omega(t)$ are unknown disturbances. Unknown disturbances include ‘unstructured uncertainties’ such as components that are not modelled, measurement noise and unknown faults. All these have been incorporated into the process model as additive faults. Observe that Eq. (4) and Eq. (6) are essentially the same except for the noise term in the latter. A comparison of Eq. (6) and Eq. (5) reveals another distinction between additive and multiplicative faults. Additive faults occur in the model as unknown functions of time multiplying known matrices, whereas multiplicative faults occur in the model as known functions of time (observable) multiplying unknown matrices. These differences have a bearing on how they are treated in the diagnostic methodology (Gertler, 1992). Additive and multiplicative faults are often referred to as uncertainties in control literature and are distinguished for the same reason (Morari & Zafiriou, 1989).

Methods based on analytical redundancy derive residuals which are insensitive to uncertainties but are sensitive to faults. One of the popular ways of doing this is the method of disturbance decoupling. In this approach, all uncertainties are treated as disturbances and filters (also known as unknown input observer (UIO)) are designed (Frank & Wünnenberg, 1989; Viswanadham & Srichander, 1987) to decouple the effect of faults and unknown inputs so that they can be differentiated. In chemical engineering systems, a *tuples* method that exploits the functional redundancy using steady state mass and energy balance equations was proposed by Kavuri and Venkatasubramanian (1992). In this method, constraints are developed for various set of assumptions using model equations and these constraints are monitored for violation. While the use of steady state model leads to a satisfactory diagnostic scheme, the use of dynamic model would reduce the modelling errors and improve the relations between the tuple deviation and the corresponding assumptions.

5.2. Residual generation in dynamic systems

The analytical redundancy schemes for fault diagnosis are basically signal processing techniques using state estimation, parameter estimation, adaptive filtering and so on. Both of above models, state–space or input–output alike, can be written as

$$\mathbf{y}(t) = f(\mathbf{u}(t), \omega(t),\ \mathbf{x}(t),\ \theta(t))$$

where $\mathbf{y}(t)$, $\mathbf{u}(t)$ denote the measurable outputs and inputs, $\mathbf{x}(t)$ and $\omega(t)$ represent (mostly unmeasurable) state variables and disturbance, and θ is the process parameters. Process faults usually cause changes in the state variables and/or changes in the model parameters. Based on the process model, one can estimate the unmeasurable $\mathbf{x}(t)$ or $\theta(t)$ by the observed $\mathbf{y}(t)$ and $\mathbf{u}(t)$ using state estimation and parameter estimation methods. Kalman filters or observers have been widely used for state estimation (Frank & Wünnenberg, 1989). Least squares methods provide a powerful tool by monitoring the parameter estimates online (Isermann, 1989). More recently, techniques relying on parity equations for residual generation have also been developed (Chow & Willsky, 1984; Gertler, Fang & Luo, 1990). Parity equations are obtained by rearranging or transforming the input–output models, which are relatively easy to generate from on-line process data and easy to use. Several most frequently used residual generation methods are discussed in this section.

5.2.1. Diagnostic observers for dynamic systems

The main concern of observer-based FDI is the generation of a set of residuals which detect and uniquely identify different faults. These residuals should be robust in the sense that the decisions are not corrupted by such unknown inputs as unstructured uncertainties like process and measurement noise and modelling uncertainties. The method develops a set of observers, each one of which is sensitive to a subset of faults while insensitive to the remaining faults and the unknown inputs. The extra degrees of freedom resulting from measurement and model redundancy make it possible to build such observers. The basic idea is that in a fault-free case, the observers track the process closely and the residuals from the unknown inputs will be small. If a fault occurs, all observers which are made insensitive to the fault by design continue to develop small residuals that only reflect the unknown inputs. On the other hand, observers which are sensitive to the fault will deviate from the process significantly and result in residuals of large magnitude. The set of observers is so designed that the residuals from these observers result in distinct residual pattern for each fault, which makes the fault isolation possible. Unique fault signature is guaranteed by design where the observers show complete fault decoupling and invariance to unknown disturbances while being independent of the fault modes and nature of disturbances.

To see how one constructs an UIO, consider a system described by the following discrete time state–space equations.

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{E}\mathbf{d}(t) + \mathbf{F}\mathbf{p}(t)$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) \tag{7}$$

where $\mathbf{d}(t)$ stands for the unknown inputs. An observer is a model that takes the form

$$\mathbf{x}_0(t) = \mathbf{T}\mathbf{x}(t)$$

$$\mathbf{x}_0(t+1) = \mathbf{H}\mathbf{x}_0(t) + \mathbf{J}\mathbf{u}(t) + \mathbf{G}\mathbf{y}(t) \tag{8}$$

The idea is to use a dynamic algorithm to estimate the state variables from observed inputs and outputs. The design of an observer amounts to the choices of matrices $\mathbf{T}$, $\mathbf{H}$, $\mathbf{J}$ and $\mathbf{G}$. Denote the estimation error (the state estimation error) at instant $t+1$ by $\mathbf{e}(t+1)$ and the residual by $\mathbf{r}(t)$. Then

$$\mathbf{e}(t+1) = \mathbf{x}_0(t+1) - \mathbf{T}\mathbf{x}(t+1)$$

$$\mathbf{r}(t) = \mathbf{L}_1\mathbf{x}_0(t) + \mathbf{L}_2\mathbf{y}(t) \tag{9}$$

It is easy to show that

$$\begin{aligned}\mathbf{e}(t+1) = {} & \mathbf{H}\mathbf{x}_0(t) + (\mathbf{J} - \mathbf{T}\mathbf{B})\mathbf{u}(t) + \mathbf{G}\mathbf{y}(t) - \mathbf{T}\mathbf{A}\mathbf{x}(t) \\ & - \mathbf{T}\mathbf{E}\mathbf{d}(t) - \mathbf{T}\mathbf{F}\mathbf{p}(t)\end{aligned} \tag{10}$$

In order to make the observer track the system independent of the unknown inputs $\mathbf{d}(t)$, we need to choose such matrix $\mathbf{T}$ that $\mathbf{TE} = 0$. Observer tracking is unaffected by the input $\mathbf{u}(t)$ if the matrix $\mathbf{J}$ is so chosen that $\mathbf{J} = \mathbf{TB}$. Substituting these conditions into Eq. (10) yields

$$\mathbf{e}(t+1) = \mathbf{H}\mathbf{x}_0(t) + (\mathbf{GC} - \mathbf{TA})\mathbf{x}(t) - \mathbf{T}\mathbf{F}\mathbf{p}(t) \tag{11}$$

Choose matrix $\mathbf{G}$ so that $\mathbf{GC} - \mathbf{TA} = -\mathbf{HT}$, $\boldsymbol{H}$ stable and $\mathbf{L}_1\mathbf{T} + \mathbf{L}_2\mathbf{C} = 0$. Then

$$\mathbf{e}(t+1) = \mathbf{He}(t) - \mathbf{TFp}(t)$$

$$\mathbf{r}(t) = \mathbf{L}_1\mathbf{e}(t) \tag{12}$$

In fault-free case, $\mathbf{p}(t) = 0$ and

$$\mathbf{e}(t+1) = \mathbf{He}(t) \tag{13}$$

If the absolute values of the eigen values of $\mathbf{H}$ are less than 1, $\mathbf{e}(t) \to 0$ as $t \to \infty$. As a result, in fault-free case the estimation error and consequently the residual will track the process regardless of the unknown inputs $\mathbf{d}(t)$ to the process. Therefore it is named the *unknown input observer*.

When a sensor fault occurs, output $\mathbf{y}(t)$ in Eq. (7) changes to

$$\mathbf{y}(t) = \mathbf{Cx}(t) + \mathbf{q}(t) \tag{14}$$

and the estimation error and residual become

$$\mathbf{e}(t+1) = \mathbf{He}(t) + \mathbf{Gq}(t)$$

$$\mathbf{r}(t) = \mathbf{L}_1\mathbf{e}(t) + \mathbf{L}_2\mathbf{q}(t) \tag{15}$$

Consequently, the error and the residual carry the 'signature' of the sensor faults. 'Signatures' corresponding to the actuator faults are reflected in estimation error and residual as seen in Eq. (12).

We have shown, briefly, the basic idea behind the generation of diagnostic observers. For a detailed discussion on general diagnostic observer design for linear systems, the reader is referred to Frank (1994). One important issue to be noted, as pointed out by Frank, is that the observer-based design does not need the application of state estimation theory, instead, only output estimators are needed which are generally realized as filters.

It is interesting to see that the concepts of isolability, rejection of modelling uncertainties and multiple fault identifiability all come together in terms of the mathematical treatment provided above. In the observer design, some of the degrees of freedom are taken up by the condition $\mathbf{TE} = 0$ for rejecting modelling uncertainties. The degrees of freedom lost depend on the size and structure of the matrix $\mathbf{E}$. The remaining degrees of freedom can be used in decoupling fault effects for isolability and multiple fault identifiability. This depends on the structure of the $\mathbf{H}$ matrix too, which specifies system characteristics for the different fault scenarios.

Some earlier work using diagnostic observers approach can be found in (Clark, 1979; Massoumnia, 1986; Frank & Wünnenberg, 1989) among others. Frank (1990) presented a solution to the fundamental problem of robust fault detection, that provides the maximum achievable robustness by decoupling the effects of faults from each other and from the effects of modelling errors. It is desirable that the performance of an FDI scheme is unaffected by conditions in the operating system which may be different from what it was originally designed. A major problem in the robustness of an FDI scheme comes from uncertainties in the physical parameters of the operating plant. It is fair to say that the more complex the system model is and the heavier the technique depends on the model, the more important the robustness becomes (Willsky, 1976). Robustness remains an important problem in fault diagnosis.

Generation of diagnostic observers for nonlinear systems have also been considered to certain extent in the literature. An elegant approach for the generation of diagnostic observers for nonlinear systems which are in the fault-affine (similar to control-affine forms discussed in control literature) form can be found in (Frank, 1990), where a continuous time model is given. Its corresponding discrete system model can be written as

$$\mathbf{x}(t+1) = \mathbf{Ax}(t) + \mathbf{B}(\mathbf{y}(t),\ \mathbf{u}(t)) + \mathbf{Ed}(t) + \mathbf{F}(\mathbf{x}(t))\mathbf{p}(t)$$

$$\mathbf{y}(t) = \mathbf{Cx}(t) + \mathbf{K}(\mathbf{x}(t))\mathbf{p}(t), \tag{16}$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{E}$, $\mathbf{F}$, $\mathbf{C}$, $\mathbf{K}$ are matrices of appropriate dimensions. This nonlinear observer design problem follows the same line as the linear observers and the observer model takes the form

$$\mathbf{x}_0(t) = \mathbf{Tx}(t)$$

$$\mathbf{x}_0(t+1) = \mathbf{Hx}_0(t) + \mathbf{J}(\mathbf{y}(t),\ \mathbf{u}(t)) + \mathbf{Gy}(t)$$

$$\mathbf{r}(t) = \mathbf{L}_1\mathbf{x}_0(t) + \mathbf{L}_2\mathbf{y}(t) \tag{17}$$

which leads to the observer design equations

$$\mathbf{TA} - \mathbf{HT} = \mathbf{GC}$$

$$\mathbf{TE} = 0$$

$$\mathbf{J}(\mathbf{y}(t), \mathbf{u}(t)) = \mathbf{TB}(\mathbf{y}(t), \mathbf{u}(t))$$

$$\mathbf{L}_1\mathbf{T} + \mathbf{L}_2\mathbf{C} = 0 \tag{18}$$

From Eq. (18) matrices $\mathbf{T}$, $\mathbf{H}$, $\mathbf{J}$, $\mathbf{G}$, $\mathbf{L}_1$, $\mathbf{L}_2$ can be determined.

There have been other researchers who have looked at the problem of nonlinear observer design for a restricted class of nonlinear systems. Design of observers for bilinear nonlinearity can be found in (Dingli, Gomm, Shields, Williams, & Disdell, 1995). Observers based on differential geometric methods for fault-affine model forms can be found in (Yang & Saif, 1995). Another approach that can be used in the design of nonlinear observers follows the work done in the development of Kalman filters for state estimation. In this approach, the observer gain matrix becomes time-variant and is the linearized gain matrix around the current operating point. These observers work well for slowly varying faults, but might have difficulty in identifying jump failures. More information on such observers can be found in (Frank, 1990).

5.2.2. Remarks on residual evaluation

Most of the work on observer design focuses on the generation of residuals for dynamic systems with satisfactory decoupling properties. Residual evaluation also plays an important role in subsequent fault detection and diagnosis. The residual evaluation part considers the trade-off between fast and reliable detection. In most of the work on observer design, simple threshold function is used in residual evaluation. Statistical classifiers can also be used in residual evaluation. A neural network approach for residual evaluation is presented in (Koppen-Seliger, Frank, & Wolff, 1995).

5.2.3. Parity relations

Parity equations are rearranged and usually transformed variants of the input–output or state–space models of the plant (Gertler, 1991; Gertler & Singer, 1990). The essence is to check the parity (consistency) of the plant models with sensor outputs (measurements) and known process inputs. Under ideal steady state operating conditions, the so-called residual or the value of the parity equations is zero. In real situations, the residuals are nonzero due to measurement and process noise, model inaccuracies, gross errors in sensors and actuators, and faults in the plant. The idea of this approach is to rearrange the model structure so as to get the best fault isolation. Dynamic parity relations was introduced by Willsky (1976). Further developments have been made by Gertler et al. (1990), Gertler, Costin, Fang, Kowalczuk, Kunwer and Monajemy (1995) and Gertler and Monajemy (1995) among others. Vaclavek (1984) suggested the use of short-term averages of the steady state balance equation residuals. Almasy and Sztano (1975) used residuals to identify gross bias faults. Redundancy provides freedom in the design of residual generating equations so that further fault isolation can be achieved. Fault isolation requires the ability to generate residual vectors which are orthogonal to each other for different faults. Ben-Haim (1980) used redundancy in the balance equations for generating orthogonal residuals for different fault classes. He also extended the approach (Ben-Haim, 1983) to dynamic systems to guarantee isolability under ideal conditions. Chow and Willsky (1984) proposed a procedure to generate parity equations from the state–space representation of a dynamic system. By defining marginal sizes for fault alarms, Gertler and Singer (1990) extended it to statistical isolability under noisy conditions and generalized the isolability criteria by simultaneously minimizing the sensitivity of the residuals to small drifts in cases having only the additive plant faults. They are attractive alternatives owing to their ability to determine, a priori, isolability of different faults. However, it should be noted that all these methods are limited to faults that do not include gross process parameter drifts, and none of them address the issue of significant uncertainties in multiplicative parametric faults.

The idea of parity space approaches can be explained as follows (Desai & Ray, 1984; Frank, 1990)). Let $y \in \mathbb{R}^n$ be the measurement vector; $\mathbf{x} \in \mathbb{R}^m$ be the true values of the state variables. Redundancy exists if $n > m$. Under fault-free conditions, let $\mathbf{y}$ and $\mathbf{x}$ be related by

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) \tag{19}$$

When a fault occurs with one of the measurements

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \Delta\mathbf{y}(t) \tag{20}$$

where $\mathbf{C} \in \mathbb{R}^{n\times m}$ is the parameter matrix. Choose the projection matrix $\mathbf{V} \in \mathbb{R}^{(n-m)\times n}$ satisfying

$$\mathbf{V}\mathbf{C} = 0$$

$$\mathbf{V}^T\mathbf{V} = \mathbf{I}_n - \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T \tag{21}$$

Being a null space of $\mathbf{C}$, the rows of $\mathbf{V}$ are required to be orthogonal, i.e. $\mathbf{V}\mathbf{V}^T = \mathbf{I}_{n-m}$. Combining the observation $\mathbf{y}$ into a parity vector $\mathbf{p}$ yields

$$\mathbf{p}(t) = \mathbf{V}\mathbf{y}(t) = \mathbf{V}\mathbf{C}\mathbf{x}(t) + \mathbf{V}\Delta\mathbf{y}(t) = \mathbf{V}\Delta\mathbf{y}(t) \tag{22}$$

$\mathbf{p}(t) = \mathbf{V}\mathbf{y}(t)$ is the parity equation set whose residuals carry the signature of the measurement faults. In the fault-free case, $\mathbf{p} = 0$. For a single i^{th} sensor fault.

$$\Delta\mathbf{y} = [0 \quad 0 \quad 0 \quad \dots \quad \Delta\mathbf{y}_i \quad \dots \quad 0]'$$

$$\mathbf{V}\Delta\mathbf{y} = \Delta\mathbf{y}_i \cdot (\text{the } i^{\text{th}} \text{ column of } \mathbf{V}). \tag{23}$$

Thus the columns of $\mathbf{V}$ determine n distinct directions associated with those n sensor faults, which enables the distinction of the n fault signatures and hence their isolability.

The above procedure assumes direct redundancy. Due to Chow and Willsky (1984), the following procedure provides a general scheme for both direct and temporal redundancy. Consider the state–space model (1). The output at $t+1$ is

$$\mathbf{y}(t+1) = \mathbf{C}\mathbf{A}\mathbf{x}(t) + \mathbf{C}\mathbf{B}\mathbf{u}(t) + \mathbf{D}\mathbf{u}(t+1) \tag{24}$$

For any $s > 0$, $\mathbf{y}(t+s)$ takes the form

$$\mathbf{y}(t+s) = \mathbf{C}\mathbf{A}^s\mathbf{x}(t) + \mathbf{C}\mathbf{A}^{s-1}\mathbf{B}\mathbf{u}(t) + \dots + \mathbf{C}\mathbf{B}\mathbf{u}(t+s-1) + \mathbf{D}\mathbf{u}(t+s) \tag{25}$$

Collecting the equations for $s = 0, 1, \dots, n_1 \le n$ and writing it in a compact form yield

$$\mathbf{Y}(t) = \mathbf{Q}\mathbf{x}(t-n_1) + \mathbf{R}\mathbf{U}(t) \tag{26}$$

Pre-multiplying Eq. (26) with a vector $\mathbf{w}^T$ of appropriate dimension yields a scalar equation

$$\mathbf{w}^T\mathbf{Y}(t) = \mathbf{w}^T\mathbf{Q}\mathbf{x}(t-n_1) + \mathbf{w}^T\mathbf{R}\mathbf{U}(t) \tag{27}$$

In general, this equation will contain input variables, output variables and unknown state variables. It will qualify as a parity equation only if the state variables disappear which requires

$$\mathbf{w}^T\mathbf{Q} = 0 \tag{28}$$

This is a set of homogeneous linear equations. If the system is observable, these m equations are independent. It has been shown that once the design objectives are selected, parity equation and observer-based designs lead to identical or equivalent residual generators (Gertler, 1991).

5.2.4. Kalman filters

The plant disturbances are random fluctuations and oftentimes only their statistical parameters are known. One solution (Basseville, 1988; Willsky, 1976) to the fault diagnosis problem in such systems entails monitoring the innovation process or the prediction errors. The objective is to design a state estimator with minimum estimation error. It involves the use of optimal state estimate, e.g. the Kalman filter, which is designed on the basis of the system model in its normal operating mode.

It is well known that the Kalman filter is a recursive algorithm for state estimation and it has found wide applications in chemical as well as other industrial processes. The Kalman filter in state–space model is equivalent to an optimal predictor for a linear stochastic system in the input–output model. The essential Kalman filter theory can be summarized briefly as follows.

Describe a linear finite dimensional stochastic system by a state–space model

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \omega(t)$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + v(t), \quad t \geq 0 \tag{29}$$

where $\mathbf{x}(t)$ is n-dimensional vector, $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are matrices with suitable dimensions, $\mathbf{x}_0$ has mean $\bar{\mathbf{x}}_0$ and covariance Σ_0; $\omega(t)$ and $v(t)$ are Gaussian white noise sequences with means $E\{\omega(t)\} = 0$, $E\{v(t)\} = 0$ and the covariance matrix

$$E\left\{ \begin{pmatrix} \omega(t) \\ v(t) \end{pmatrix} (\omega^T(\tau), v^T(\tau)) \right\} = \begin{pmatrix} \mathbf{Q} & \mathbf{S} \\ \mathbf{S}' & \mathbf{R} \end{pmatrix} \delta_{t-\tau}, \tag{30}$$

where $\delta_{t-\tau}$ is Kronecker's delta, $\omega(t)$, $v(t)$ are independent of $\sigma(x_s{:}s \leq t)$.

In estimating the state $\mathbf{x}(t+1)$ based on the observed data $\{\mathbf{y}(t)\}$ and $\{\mathbf{u}(\mathrm{t})\}$, the optimal Kalman filter minimizes the function

$$J = \lim_{t \to \infty} E(\mathbf{e}^T(t)\mathbf{e}(t)) \tag{31}$$

where $\mathbf{e}(t)$ is the estimation error and is defined as $\mathbf{e}(t) = \mathbf{y}(t) - \mathrm{C}\hat{\mathbf{x}}(t)$.

Assume the initial state and noise sequences are jointly Gaussian. Consider the estimator $\hat{\mathbf{x}}(t+1) = E\{\mathbf{x}(t+1)|\mathbf{y}(t), \ldots, \mathbf{y}(0), \mathbf{u}(t), \ldots, \mathbf{u}(0)\}$. The filtered state $\hat{\mathbf{x}}(t+1)$ satisfies:

$$\hat{\mathbf{x}}(t+1) = \mathbf{A}\hat{\mathbf{x}}(t) + \mathbf{B}\mathbf{u}(t) + K(t)[\mathbf{y}(t) - \mathbf{C}\hat{\mathbf{x}}(t)]$$

$$\hat{\mathbf{x}}_0 = \bar{\mathbf{x}}_0. \tag{32}$$

The Kalman filter gain $K(t)$ is given by

$$K(t) = [\mathbf{A}\Sigma(t)\mathbf{C}^T + \mathbf{S}][\mathbf{C}\Sigma(t)\mathbf{C}^T + \mathbf{R}]^{-1} \tag{33}$$

where $\Sigma(t)$ is a $n \times n$ state error covariance matrix.

The statistical analysis of Kalman filter was pioneered by Willsky and Jones (1976) and further explored by Basseville and Benveniste (1986) and Basseville and Nikiforov (1993) and the references therein. It has been shown that a bank of Kalman filters (Basseville & Benveniste) designed on the basis of all the available possible system models under all possible changes can be used for the isolation purpose. Fathi, Ramirez, and Korbiez (1993) included adaptive analytical redundancy models in the diagnostic reasoning loop of knowledge-based systems. The modified extended Kalman filter (EKF) is used in designing local detection filters in their work. In a recent work, Chang and Hwang (1998) explored the possibility of using sub optimal EKF in order to enhance computation efficiency without sacrificing diagnosis accuracy.

5.2.5. Parameter estimation

Diagnosis of parameter drifts which are not measurable directly requires on-line parameter estimation methods. Accurate parametric models of the process are needed, usually in the continuous domain in the form of ordinary and partial differential equations. The system models represented by Eq. (1) and Eq. (7) assume process parameters to be either constants or dependent only on state variables. Faults which occur as time dependent parameter drifts can be handled through parameter estimation methods. This procedure (Isermann, 1984) is described as follows: obtain process model with only the measured inputs and outputs in the form:

$$\mathbf{y}(t) = f(\mathbf{u}(t), \theta). \tag{34}$$

Model parameters θ are estimated as measurements $\mathbf{y}(t)$ and $\mathbf{u}(t)$ become available, θ in turn are related to physical parameters φ in the process by $\theta = g(\varphi)$. Changes in the parameters φ, $\Delta\varphi$ are computed from this relationship. Using methods of pattern recognition, one can relate the changes $\Delta\varphi$ to process faults.

Isermann (1984) and Young (1981) surveyed different parameter estimation techniques such as least squares, instrumental variables and estimation via discrete-time models. These methods require the availability of accurate dynamic models of the process and are computationally very intensive for large processes. The most important issue in the use of parameter estimation approach for fault diagnosis is one of complexity. The process model used could be either based on input output data, nonlinear first principles model or a reduced order model. If the process model is complex nonlinear first principles model, then the parameter estimation problem turns out to be a nonlinear optimization problem. Real-time solution to complex nonlinear optimization problems is a serious bottleneck in the application of such approaches. Reduced order or input–output data models could be used in the parameter estimation approach and in this case, the robustness of the approach has to be addressed.

5.3. Hardware redundancy and voting schemes

Voting techniques are often used in systems that possess high degrees of parallel hardware redundancy (Willsky, 1976). For example, consider three identical sensors measuring the same variable. If one of the three signals differs markedly from the other two, the differing signal is identified as faulty. The difference between the two signals in every pair of sensors in a redundant group indicates a fault. Voting schemes are easy to implement and are quick to identify mechanical faults in instruments. An alternative approach (Desai and Ray, 1984) is to search, given error bounds on different sensors, for subsets of sensor measurements with different degrees of consistency. The most consistent subset is used for estimating the measured quantity. The least consistent subset is used for isolating the faulty sensors. Broen (1974) has developed a class of voter-estimators using parallel hardware redundancy to detect sensor faults. The advantage is that the faulty sensors are removed smoothly from consideration reducing the number of false alarms. Voting systems in isolation do not take advantage of singly or even doubly redundant sensors and hence ignore potential information.

5.4. Generating enhanced residuals

The diagnostic residuals reflect the potential faults of a system. The next step is to confirm the presence of a fault and to identify it, i.e. to detect and to isolate the fault. For the purpose of isolation, it is necessary to generate such diagnostic residuals that are not only fault sensitive but also fault-selective. To this end, the residual generator must be able to produce a set of residuals rather than a single one and make residuals respond selectively to each potential fault. The residuals generated can thus serve not only as fault detector but also as fault classifier.

There have been efforts to design residual generators capable of generating 'enhanced' residuals which are conducive to fault isolation. Two of such enhancement methods, the directional residual approach (Gertler and Monajemy, 1995) and the structural residual approach (Gertler et al., 1990; Gertler and Singer, 1990) have attracted much attention owing to their capability of generating residuals having directional or structural properties thereby facilitating the fault isolation process. The structured residual generators are designed in such a way that each residual responds to a subset of faults selectively. Such design allows to form binary fault signatures for further isolation. The directional residual generators are capable of generating residuals that are confined to a fault specific direction in the multidimensional residual space. As a result, the fault isolation step amounts to the determination of a predefined direction to which the residuals lie the closest.

5.4.1. Directional residuals

The directional residual approach generates residual vectors which are confined to a fault specific direction that allows to isolate the faults from their locations in the multidimensional residual space. The design of a directional residual generator is based on linear time-invariant finite dimensional systems. Consider the system model (Yin, 1998):

$$\mathbf{h}(z^{-1})\mathbf{y}(t) = \mathbf{U}(z^{-1})\mathbf{u}(t) + \mathbf{V}(z^{-1})\mathbf{p}(t) + \mathbf{W}(z^{-1})\omega(t), \quad t \geq 1 \tag{35}$$

where $\mathbf{p}(t)$, $\omega(t)$ denote faults and noise, respectively.

A residual generator is a linear dynamic operator operating on the observable $\mathbf{y}(t)$ and $\mathbf{u}(t)$ and having a form

$$\mathbf{r}(t) = \mathbf{G}(z^{-1})\mathbf{y}(t) + \mathbf{H}(z^{-1})\mathbf{u}(t) \tag{36}$$

Eq. (36) is the 'computational form' of the generator. It was so designed that $\mathbf{G}(z^{-1})$ and $\mathbf{H}(z^{-1})$ are polynomials. Such a generator is computationally simple and is guaranteed to yield bounded residuals provided that $\mathbf{y}(t)$ and $\mathbf{u}(t)$ are bounded. The residuals should not be affected by the system input $\mathbf{u}(t)$, which leads to the fundamental property of the generator

$$\mathbf{H}(z^{-1}) = -\mathbf{h}^{-1}(z^{-1})\mathbf{G}(z^{-1})\mathbf{U}(z^{-1}) \tag{37}$$

Combining Eqs. (35)–(37) yields

$$\mathbf{r}(t) = \mathbf{F}(z^{-1})\mathbf{p}(t) + \mathbf{L}(z^{-1})\omega(t) \tag{38}$$

Eq. (38) is the 'internal form' of a residual generator, which explains the sources of the residual.

The response of the directional residual vector $\mathbf{r}(t)$ to the combined effects of all faults and noise is

$$\mathbf{r}(t) = \Psi\Delta(z^{-1})\mathbf{p}(t) + \Pi\mathrm{M}(z^{-1})\omega(t) \tag{39}$$

where $\Delta(z^{-1})$ and $\mathrm{M}(z^{-1})$ describe the dynamics of the fault and the noise, respectively; and the matrices Ψ and Π govern the directions of the fault and the noise.

Gertler and Monajemy (1995) have shown that the directional residuals in response to an arbitrary mix of input and output faults can be generated by using dynamic parity relations as well as by observer-based designs. Their design relies on the input–output model of the monitored system, and the parity relations are directly applied to the measurements of the input and output. A classification procedure which uses multivariate statistical anlysis to minimize the expected loss from misclassification was designed in Yin and Gertler (1995). To balance between optimality and robustness, a minimax procedure (Yin, 1998) was proposed to deal with more general situation.

5.4.2. Structured residuals

The structured residual approach produces residual vectors so that each residual element, responds selectively to a subset of faults. It is required that only a fault-specific subset of the residual components is nonzero in response to a fault. Equivalently, the residuals corresponding to a specific fault will be confined to a subspace of the entire residual space. That allows to form binary fault signatures, or the so-called residual structure for fault identification (isolation) since each residual is completely unaffected by a different subset of faults.

The structured residuals may be generated by structured parity equations in either ARMA or MA format, or by state–space equations. They can also be generated using the direct eigenstructure assignment of the diagnostic observer. The following briefly describes how the diagnostic residuals with structured properties can be generated.

For a linear system, the observed input $\mathbf{u}(t)$ and output $\mathbf{y}(t)$ are related to their 'true' values $\mathbf{u}^0(t)$ and $\mathbf{y}^0(t)$ by

$$\mathbf{u}(t) = \mathbf{u}^0(t) + \mathbf{p}(t)$$

$$\mathbf{y}(t) = \mathbf{y}^0(t) + \mathbf{q}(t)$$

where $\mathbf{p}(t)$ and $\mathbf{q}(t)$ stand for actuator and sensor faults, respectively. A residual is simply defined as

$$\mathbf{o}(t) = \mathbf{H}(z)\mathbf{y}(t) - \mathbf{G}(z)\mathbf{u}(t) \tag{40}$$

Rewrite Eq. (40) as

$$\mathbf{o}(t) = \mathbf{H}(z)(\mathbf{y}^0(t) + \mathbf{q}(t)) - \mathbf{G}(z)((\mathbf{u}^0(t) + \mathbf{p}(t)) \tag{41}$$

Equivalently (from Eq. (2)):

$$\mathbf{o}(t) = \mathbf{H}(z)\mathbf{q}(t) - \mathbf{G}(z)\mathbf{p}(t) \tag{42}$$

The diagnosis residuals $\mathbf{r}(t)$ having desirable structural properties are achievable with further transformations

$$\mathbf{r}(t) = \mathbf{W}(z)\mathbf{o}(t) \tag{43}$$

Observe that a proper choice of matrix $\mathbf{W}(z)$ will allow us to implement certain properties on $\mathbf{r}(t)$.

The residual structures are characterized by incidence matrices, whose columns and rows are fault codes and residuals, respectively. For a system with three possible faults $\mathbf{F} = [F_1\ F_2\ F_3]'$, for example, a possible incidence matrix is (Gertler et al., 1990)

$$\begin{pmatrix} & F_1 & F_2 & F_3 \\ r_1 & I & I & 0 \\ r_2 & 0 & I & I \\ r_3 & I & 0 & I \end{pmatrix} \tag{44}$$

where an I element indicates that the residual does respond to the fault while a 0 means that it does not. Columns of the incidence matrix are the signatures of the particular fault. Therefore a fault is not detectable if its corresponding column in the incidence matrix contains all zeros, which means that no residuals are responsive to the fault. Two faults are not distinguishable by a structure if their columns are identical.

5.5. Further discussion on quantitative model-based diagnosis methods

It can be seen that one of the major advantages of using the quantitative model-based approach is that we will have some control over the behavior of the residuals. However, several factors such as system complexity, high dimensionality, process nonlinearity and/or lack of good data often render it very difficult even impractical, to develop an accurate mathematical model for the system. This, of course, limits the usefulness of this approach in real industrial processes.

The evaluation of residuals usually involves threshold testing. Statistical tests have been utilized for residuals generated from parity relation as well as observer-based designs. The diagnostic residuals are correlated, that has complicated the test design and execution procedures. The correlation issue was addressed in (Gertler & Yin, 1996). To better handle the correlation issue, probability distributions of the correlated residuals as well as their moving averages were derived. The Generalized Likelihood Ratio test was applied to correlated residuals for detection and isolation.

Many survey papers with different emphases on various model-based approaches have been published over the past two decades. The earliest was due to Willsky (1976), that covers methods ranging from the design of specific fault-sensitive filters to the use of statistical tests on filter innovations and the development of jump process formulations. The issue of complexity vs. performance was also addressed therein. Isermann (1984) reviewed the fault detection methods based on the estimation of non-measurable process parameters and state variables. Basseville (1988) addressed the problem of detection, estimation and diagnosis of changes in dynamic properties of signals or systems with emphasis on statistical methods. Frank (1990) outlined the principles and the most important techniques using parameter identification and state estimation with emphasis on the robustness with respect to modelling errors. Gertler (1991) presented several residual generation methods including parity equations, diagnostic observers and Kalman filter in a consistent framework. He showed that once the desired residual properties have been selected, parity equation and observer based design lead to identical and equivalent residual generators. Frank et al. (2000), recently, have reviewed the state-of-the-art developments in model-based fault diagnosis in technical processes. Among the books published in this area, the first one addressed the issue of fault detection and diagnosis in chemical industry and was written by Himmelblau (1978). The first multi-authored book on fault diagnosis in dynamic systems (Patton, Frank, & Clark, 1989), published in 1989, provided a wide coverage of various methods. Theory and application of detection of abrupt changes can be found in Basseville and Benveniste (1986) and Basseville and Nikiforov (1993). Recent books on fault detection and diagnosis have been written by Gertler (1998) and Chen and Patton (1999) and Russell, Chiang, and Braatz (2000). A number of other references can be found in the survey papers and books cited above.

Most of the work on model-based diagnostic systems has so far been concentrated in the aerospace, mechanical and electrical engineering literature. There has not been too much work on applications of observers for fault diagnosis in chemical engineering systems. However, the idea and use of state estimators is quite prevalent in the process control community (Soroush, 1998). This may be due to the fact that the objectives of state/parameter estimation techniques when used in process control are different from the objectives of fault diagnosis as pointed out by Watanabe and Himmelblau (1984). Further, the unavailability/complexity of high fidelity models and the essential nonlinear nature of these models for chemical processes render the design of diagnostic observers for chemical engineering systems quite difficult. Watanabe and Himmelblau (1982) looked at the detection of instrument faults in nonlinear time-varying processes that include uncertainties such as modelling error, parameter ambiguity and input/output noise using state estimation filters. The same authors later (Watanabe and Himmelblau, 1983a,b) proposed a two-level strategy for fault detection and diagnosis. The first-level consists of estimation of the states of a nonlinear process by a linear state estimator (Luenberger observer). The second level consisted of fault parameter identification by using the estimated state vector from the first-level through a least squares procedure. This technique was applied to a chemical reactor and shown to be faster that EKF as well as yielding unbiased and precise estimates. Watanabe and Himmelblau (1984) used a linear reduced order state estimator and the EKF to reconstruct state variables and to identify the unknown deteriorating parameters in the same reactor. They show this to be more effective that using just the EKF on the same process model.

Gertler and Luo (1989) presented a design procedure to generate isolable parity equation models chosen from a multitude of suitable models on the basis of sensitivities with respect to different failures and robustness relative to uncertainties in selected parameters. They illustrated the application of the technique on a distillation column. The application of EKF-based FDI system for fault diagnosis in the Model IV FCCU case study involving DAEs was reported by Huang, Dash, Reklaitis, and Venkatasubramanian (2000). The application of three UIOs, one a *linear*, second an *extended linear* and the third a *non-linear* UIO on a CSTR case study is discussed in Dash, Kantharao, Rengaswamy, and Venkatasubramanian (2001). In their work, the performance of these three observers are evaluated through extensive simulation studies.

6. Conclusions

In this first part of the three part review paper, we have reviewed quantitative model based approaches to fault diagnosis. For the comparative evaluation of various fault diagnosis methods, we first proposed a set of desirable characteristics that one would like the diagnostic systems to possess. This can serve as a common set of criteria against which the different techniques may be evaluated and compared. Further, we provided a general framework for analyzing and understanding various diagnostic systems based on the transformations of the process data before final diagnosis is performed. For quantitative model based diagnosis methods, we discussed and reviewed various issues involved in the design of fault diagnosis systems using analytical models.

In terms of the transformations in measurement space, in quantitative model based approaches, the residual equations act as the feature space. As an example, if decoupled observers can be designed for a system, then the residuals act as features that are sensitive to only one fault at a time. Residual evaluation using threshold logic maps the feature space to class space. Transformation from the decision space to class space depends on the type of observer design. If the faults are all completely decoupled, then class space and decision space are the same. Such a view of quantitative model based approach decouples the task of observer generation from that of observer evaluation and disparate techniques can be used to solve these issues in observer design.

Quantitative model based approaches can be evaluated based on the ten desirable characteristics required of a diagnostic classifier. These address important issues, both theoretical and practical, such as how quickly the system detects a fault, can it diagnose the fault with minimum misclassification, its robustness to noise and uncertainties, adaptability, explanation facility modelling effort, computational requirements and so on.

The type of models the analytical approaches can handle are limited to linear, and in some cases, to very specific nonlinear models. For a general nonlinear model, linear approximations can prove to be poor and hence the effectiveness of these methods might be greatly reduced. Another problem in this approach is the simplistic approximation of the disturbances that include modelling errors. In most cases, the disturbance matrix includes only additive uncertainty. However, in practice, severe modelling uncertainties occurring due to parameter drifts come in the form of multiplicative uncertainties. This is a general limitation of all the model-based approaches that have been developed so far. In addition to difficulties related to modelling they do not support an explanation facility owing to their procedural nature. Further, a priori estimation of classification errors can not also be provided using these methods. Another disadvantage with these methods is that if a fault is not specifically modelled (novelty identifiability), there is no guarantee that the residuals will be able to detect it. Adaptability of these approaches to varying process conditions has also not been considered. When a large-scale process is considered, the size of the bank of filters can be very large increasing the computational complexity, though, with the recent increase in computational power and the essential linear nature of these problems, this might not be a serious bottle-neck.

Given the extensive literature on quantitative model based methods, it is difficult to include all of them in a review. However, we believe that the papers and books cited in the review would serve the reader as good resources for entry and further study. In the remaining two parts of this series, we will review qualitative model based methods and process history based techniques and conclude with a comparison of all three groups of methods.

References

Almasy, G. A., & Sztano, T. (1975). Checking and correction of measurements on the basis of linear system model. *Problems of Control and Information Theory 4*, 57–69.

Bailey, S. J. (1984). From desktop to plant floor, a CRT is the control operators window on the process. *Control Engineering 31* (6), 86–90.

Basseville, M. (1988). Detecting changes in signals and systems—a survey. *Automatica 24* (3), 309–326.

Basseville, M., & Benveniste, A. (1986). *Detection of abrupt changes in signals and dynamic systems* (Lecture Notes in Control and Information Sciences: 77). Berlin: Springer-Verlag.

Basseville, M., & Nikiforov, I. V. (1993). *Detection of abrupt changes—theory and application* (Information and System Sciences Series). Prentice Hall.

Ben-Haim, Y. (1980). An algorithm for failure location in a complex network. *Nuclear Science and Engineering 75*, 191–199.

Ben-Haim, Y. (1983). Malfunction location in linear stochastic systems-application to nuclear power plants. *Nuclear Science and Engineering 85*, 156–166.

Broen, R. B. (1974). A nonlinear voter-estimator for redundant systems. In *Proceedings of IEEE conference on decision control* (pp. 743–748), Phoenix, Arizona.

Bureau of Labor Statistics (1998). *Occupational injuries and illnesses in the united states by industry*. Washington, DC: Government Printing Office.

Chang, C. T., & Hwang, J. I. (1998). Simplification techniques for EKF computations in fault diagnosis - suboptimal gains. *Chemical Engineering Science 53* (22), 3853–3862.

Chen, J., & Patton, R. J. (1999). *Robust model-based fault diagnosis for dynamic systems*. Massachusetts: Kluwer Academic Publishers.

Chow, E. Y., & Willsky, A. S. (1984). Analytical redundancy and the design of robust failure detection systems. *IEEE Transactions on Automatic Control 29* (7), 603–614.

Clark, R. N. (1979). The dedicated observer approach to instrument fault detection. In *Proceedings of the 15th IEEE-CDC* (pp. 237–241), Ford Lauderdale, FL.

Dash, S., Kantharao, S., Rengaswamy, R., & Venkatasubramanian, V. (2001) Application and evaluation of a linear/restricted nonlinear observer to a nonlinear CSTR. In *ESCAPE*, pp. 853–858, Kolding, Denmark.

Desai, M., & Ray, A. (1984). A fault detection and isolation methodology-theory and application. In *Proceedings of American control conference* (pp. 262–270) San Diego.

Dingli, Y., Gomm, J. B., Shields, D. N., Williams, D., & Disdell, K (1995). Fault diagnosis for a gas-fired furnace using bilinear observer method. In *Proceedings of the American control conference*, Seattle, Washington.

Fathi, Z., Ramirez, W. F., & Korbiez, J. (1993). Analytical and knowledge-based redundancy for fault diagnosis in process plants. *AIChE J. 39*, 42–56.

Frank, P. M. (1990). Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy—a survey and some new results. *Automatica 26*, 459–474.

Frank, P. M. (1994). On-line fault detection in uncertain nonlinear systems using diagnostic observers: a survey. *International Journal Systems Science 25* (12), 2129–2154.

Frank, P. M., Ding, S. X., & Marcu, T. (2000). Model-based fault diagnosis in technical processes. *Transactions of the Institution of Measurement and Control 22* (1), 57–101.

Frank, P. M., & Wünnenberg, J. (1989). Robust fault diagnosis using unknown input observer schemes. In R. J. Patton, P. M. Frank & R. N. Clark (Eds.), *Fault diagnosis in dynamic systems: theory and applications*. NY: Prentice Hall.

Fukunaga, K. (1972). *Introduction to statistical pattern recognition*. New York: Academic press.

Gertler, J., (1991). Analytical redundancy methods in fault detection and isolation. In *Proceedings of IFAC/IAMCS symposium on safe process* (p. 91), Baden-Baden.

Gertler, J., (1992). Analytical redundancy methods in fault detection and isolation-survey and synthesis. In *IFAC symposium on online fault detection and supervision in the chemical process industries*.

Gertler, J. (1993). Residual generation in model-based fault diagnosis. *Control-theory and advanced technology 9* (1), 259–285.

Gertler, J. (1998). *Fault detection and diagnosis in engineering systems*. Marcel Dekker.

Gertler, J., Costin, M., Fang, X., Kowalczuk, Z., Kunwer, M., & Monajemy, R. (1995). Model based diagnosis for automative engines - algorithm development and testing on a production vehicle. *IEEE Transactions on Control Systems Technology 3*, 61–69.

Gertler, J., Fang, X., & Luo, Q. (1990). Detection and diagnosis of plant failures: the orthogonal parity equation approach. *Control and Dynamic Systems 37*, 159–216.

Gertler, J., & Luo, Q. (1989). Robust isolable models for failure diagnosis. *AIChE 31* (11), 1856–1868.

Gertler, J., & Monajemy, R. (1995). Generating directional residuals with dynamic parity relations. *Automatica 31*, 627–635.

Gertler, J., & Singer, D. (1990). A new structural framework for parity equation-based failure detection and isolation. *Automatica 26*, 381–388.

Gertler, J., & Yin, K., (1996). Statistical decision making for dynamic parity relations. In *Preprint IFAC'96 13th World congress*, San Francisco.

Himmelblau, D. M. (1978). *Fault detection and diagnosis in chemical and petrochemical processes*. Amsterdam: Elsevier press.

Hoskins, J. C., & Himmelblau, D. M. (1988). Artificial neural network models of knowledge representation in chemical engineering. *Computers and Chemical Engineering 12*, 881–890.

Huang, Y., Dash, S., Reklaitis, G.V., & Venkatasubramanian, V., (2000). EKF based estimator for FDI in Model IV FCCU. *IFAC Proceedings SAFEPROCESS2000* Budapest, Hungary.

Isermann, R. (1984). Process fauls detection based on modelling and estimation methods—a survey. *Automatica 20* (4), 387–404.

Isermann, R. (1989). Process fault diagnosis based on dynamic models and parameter estimation methods. In R. J. Patton, P. M. Frank & R. N. Clark (Eds.), *Fault diagnosis in dynamic systems: theory and applications*. NY: Prentice Hall.

Kavuri, S. N., & Venkatasubramanian, V. (1992). Combining pattern classification and assumption-based techniques for process fault diagnosis. *Computers and Chemical Engineering 16* (4), 299–312.

Koppen-Seliger, B., Frank, P.M., & Wolff A., (1995). Residual evaluation for fault detection and isolation with RCE neural networks. In *Proceedings of the American control conference*, Seattle, Washington, pp. 3264–3268.

Kramer, M. A. (1987). Malfunction diagnosis using quantitative models with non-boolean reasoning in expert systems. *AIChE J. 33* (1), 130–140.

Kramer, M. A., & Mah, R. S. H., (1993). Model-based monitoring. In: Rippin, D., Hale, J., & Davis, J., (Eds.), *Proceedings of the second international conference on 'foundations of computer-aided process operations' (FOCAPO)* (pp. 45–68).

Laser, M. (2000). Recent safety and environmental legislation. *Trans IchemE 78 (B)*, 419–422.

Lees, F. P. (1996). *Loss prevention in process industries: hazard identification, assessment and control*. London: Butterworth-Heinemann.

Massoumnia, M. A. (1986). A geometric approach to the synthesis of failure detection filters. *IEEE Transactions on Automatic Control AC-31*, 839–846.

McGraw-Hill Economics (1985). *Survey of investment in employee safety and health*. NY: McGraw-Hill Publishing Co.

Mehra, R. K., & Peschon, J. (1971). An innovations approach to fault detection and diagnosis in dynamic systems. *Automatica 7*, 637–640.

Milne, R. (1987). Strategies for diagnosis. *IEEE Transactions on System, Man and Cyber 17* (3), 333–339.

Morari, M., & Zafiriou, E. (1989). *Robust process control*. New Jersey: Prentice Hall, Englewood Cliffs.

National Safety Council (1999). *Injury facts 1999 Edition*, Chicago: National Safety Council.

Nimmo, I. (1995). Adequately address abnormal situation operations. *Chemical Engineering Progress 91* (9), 36–45.

Patton, R., Frank, P., & Clark, R. (1989). *Fault Diagnosis in Dynamic Systems: Theory and Applications*. New York: Prentice Hall.

Russell, E. L., Chiang, L. H., & Braatz, R. D. (2000). *Data-driven techniques for fault detection and diagnosis in chemical processes*. London: Springer.

Soroush, M. (1998). State and parameter estimations and their applications in process control. *Computers and Chemical Engineering 23*, 229–245.

Vaclavek, V., (1984). Gross systematic errors or biases in the balance calculations. In *Papers of the Prague Institute of technology*, Prague, Czechoslovakia.

Viswanadham, N., & Srichander, R. (1987). Fault detection using unknown-input observers. *Control Theory and Advanced Technology 3*, 91–101.

Watanabe, K., & Himmelblau, D. M. (1982). Instrument fault detection in systems with uncertainties. *International Journal System Science 13* (2), 137–158.

Watanabe, K., & Himmelblau, D. M. (1983a). Fault diagnosis in nonlinear chemical processes—part I. Theory. *AIChE 29* (2), 243–249.

Watanabe, K., & Himmelblau, D. M. (1983b). Fault diagnosis in nonlinear chemical processes—part II. Application to a chemical reactor. *AIChE 29* (2), 250–260.

Watanabe, K., & Himmelblau, D. M. (1984). Incipient fault diagnosis of nonlinear processes with multiple causes of faults. *Chemical Engineering Science 39* (3), 491–508.

Willsky, A. S. (1976). A survey of design methods for failure detection in dynamic systems. *Automatica 12*, 601–611.

Willsky, A. S., & Jones, H. L. (1976). A generlized likelihood ratio approach to the detection and estimation of jumps in linear systems. *IEEE Transactions on Automatic Control AC-21*, 108–112.

Yang, H. & Saif, M., (1995). Nonlinear adaptive observer design for fault detection. In *Proceedings of the American Control Conference*, Seattle, Washington, pp. 1136–1139.

Yin, K. (1998). Minimax methods for fault isolation in the directional residual approach. *Chemical Engineering Science 53*, 2921–2931.

Yin, K., & Gertler, J., (1995). Fault detection and isolation in the directional residual approach. In *Preprint IFAC workshop on on-line fault detection and supervision in the chemical and process industries* (pp. 194–199). Newcastle, UK.

Young, P. (1981). Parameter estimation for continuous time models-a survey. *Automatica 17* (1), 23–39.